

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:54:39 ; Search time 7111.01 Seconds  
(without alignments)  
668.249 Million cell updates/sec

Title: US-09-101-423A-4  
Perfect score: 1088  
Sequence: 1 agtgggacacagcttgct.....acataataattttctgc 1088

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1033670 seqs, 2183789903 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl :

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_p11.\*
- 8: gb\_p12.\*
- 9: gb\_p13.\*
- 10: gb\_p14.\*
- 11: gb\_p15.\*
- 12: gb\_p16.\*
- 13: gb\_p17.\*
- 14: gb\_p18.\*
- 15: gb\_p19.\*
- 16: gb\_p20.\*
- 17: gb\_p21.\*
- 18: gb\_p22.\*
- 19: gb\_p23.\*
- 20: gb\_p24.\*
- 21: gb\_p25.\*
- 22: gb\_p26.\*
- 23: gb\_p27.\*
- 24: gb\_p28.\*
- 25: gb\_p29.\*
- 26: gb\_p30.\*
- 27: gb\_p31.\*
- 28: gb\_p32.\*
- 29: gb\_p33.\*
- 30: gb\_p34.\*
- 31: gb\_p35.\*
- 32: gb\_p36.\*
- 33: gb\_p37.\*
- 34: gb\_p38.\*
- 35: gb\_p39.\*
- 36: gb\_p40.\*
- 37: gb\_p41.\*
- 38: gb\_p42.\*
- 39: gb\_p43.\*
- 40: gb\_p44.\*
- 41: gb\_p45.\*
- 42: gb\_p46.\*
- 43: gb\_p47.\*

- 44: em\_hum5.\*
- 45: gb\_p13.\*
- 46: gb\_p15.\*
- 47: gb\_p16.\*
- 48: gb\_p17.\*
- 49: gb\_p18.\*
- 50: gb\_p19.\*
- 51: gb\_p20.\*
- 52: gb\_p21.\*
- 53: gb\_p22.\*
- 54: gb\_p23.\*
- 55: gb\_p24.\*
- 56: gb\_p25.\*
- 57: gb\_p26.\*
- 58: gb\_p27.\*
- 59: gb\_p28.\*
- 60: gb\_p29.\*
- 61: gb\_p30.\*
- 62: gb\_p31.\*
- 63: gb\_p32.\*
- 64: gb\_p33.\*
- 65: gb\_p34.\*
- 66: gb\_p35.\*
- 67: gb\_p36.\*
- 68: gb\_p37.\*
- 69: gb\_p38.\*
- 70: gb\_p39.\*
- 71: gb\_p40.\*
- 72: gb\_p41.\*
- 73: gb\_p42.\*
- 74: gb\_p43.\*
- 75: gb\_p44.\*
- 76: gb\_p45.\*
- 77: gb\_p46.\*
- 78: gb\_p47.\*
- 79: gb\_p48.\*
- 80: gb\_p49.\*
- 81: gb\_p50.\*
- 82: gb\_p51.\*
- 83: gb\_p52.\*
- 84: gb\_p53.\*
- 85: gb\_p54.\*
- 86: gb\_p55.\*
- 87: gb\_p56.\*
- 88: gb\_p57.\*
- 89: gb\_p58.\*
- 90: gb\_p59.\*
- 91: gb\_p60.\*
- 92: gb\_p61.\*
- 93: gb\_p62.\*
- 94: gb\_p63.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1088	100.0	1088	5	A64061
2	23	2.1	162730	51	AC026054
3	22	2.0	172219	87	CNS06C72
4	21	1.9	61344	57	AC073443
5	21	1.9	149605	9	AC005522
6	21	1.9	153422	49	AC023408
7	21	1.9	155655	49	AC023851
8	21	1.9	155849	84	AL355812
9	21	1.9	163300	49	AC023369
10	21	1.9	166484	51	AC026830
11	21	1.9	171319	64	AL162253
12	21	1.9	171370	9	AC004021

- A64061 Sequence 4
- AC026054 Homo sapi
- AL391153 Homo sapi
- AC073443 Homo sapi
- AC005522 Homo sapi
- AC023408 Homo sapi
- AC023851 Homo sapi
- AL355812 Homo sapi
- AC023369 Homo sapi
- AC026830 Homo sapi
- AL162253 Homo sapi
- AC004021 Human PAC





	PROGRESS ***	26 unordered pieces.
ACCESSION	AC026054	
VERSION	AC026054.3	GI:9309523
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 162730)	
AUTHORS	Waterston,R.H.	
TITLE	The sequence of Homo sapiens clone	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 162730)	
AUTHORS	Waterston,R.H.	
TITLE	Direct Submission	
JOURNAL	Submitted (19-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
COMMENT	On Jul 21, 2000 this sequence version replaced gi:7343219.	
	----- Genome Center -----	
	Center: Washington University Genome Sequencing Center	
	Center code: WUGSC	
	Web site:http://genome.wustl.edu/gsc/index.shtml	
	----- Project Information -----	
	Center project name: H_NH0131015	
	----- Summary Statistics -----	
	Sequencing vector: M13; 100%	
	Chemistry: Dye-primer; Er; 100% of reads	
	Assembly: Dye-terminator; Big dye; 0% of reads	
	Consensus quality: 147798 bases at least Q40	
	Consensus quality: 152556 bases at least Q30	
	Consensus quality: 154714 bases at least Q20	
	Insert size: 187000; agarose-fp	
	Quality coverage: 4.20 in Q20 bases; agarose-fp	
	Quality coverage: 4.43 in Q20 bases; sum-of-contigs	
	-----	
	* NOTE: This is a 'working draft' sequence. It currently	
	* consists of 26 contigs. The true order of the pieces	
	* is not known and their order in this sequence record is	
	* arbitrary. Gaps between the contigs are represented as	
	* runs of N, but the exact sizes of the gaps are unknown.	
	* This record will be updated with the finished sequence	
	* as soon as it is available and the accession number will	
	* be preserved.	
	1	
	2845	2844: contig of 2844 bp in length
	2945	2944: gap of unknown length
	4474	4473: contig of 1529 bp in length
	4574	4573: gap of unknown length
	6040	6039: contig of 1466 bp in length
	6140	6139: gap of unknown length
	7693	7692: contig of 1553 bp in length
	7793	7792: gap of unknown length
	10701	10700: contig of 2908 bp in length
	10801	10800: gap of unknown length
	13658	13657: contig of 2857 bp in length
	13758	13757: gap of unknown length
	17677	17676: contig of 3919 bp in length
	17777	17776: gap of unknown length
	22692	22691: contig of 4915 bp in length
	22792	22791: gap of unknown length
	26420	26419: contig of 3628 bp in length
	26520	26519: gap of unknown length
	29392	29391: contig of 2872 bp in length
	29492	29491: gap of unknown length
	32372	32371: contig of 3780 bp in length
	33272	33271: gap of unknown length
	33372	33371: contig of 4885 bp in length
	38257	38256: gap of unknown length
	38357	38356: contig of 4135 bp in length
	*	
	42591	gap of unknown length
	47107	contig of 4516 bp in length
	47208	gap of unknown length
	51071	contig of 3864 bp in length
	51171	gap of unknown length
	57656	contig of 6485 bp in length
	57757	gap of unknown length
	58345	contig of 7589 bp in length
	65346	gap of unknown length
	72867	contig of 7421 bp in length
	72966	gap of unknown length
	82352	contig of 9386 bp in length
	82452	gap of unknown length
	90681	contig of 8229 bp in length
	90782	contig of 9283 bp in length
	100064	contig of unknown length
	100164	gap of unknown length
	112323	contig of 12159 bp in length
	112423	gap of unknown length
	122332	contig of 9909 bp in length
	122433	gap of unknown length
	133727	contig of 11295 bp in length
	133827	gap of unknown length
	133828	contig of 14253 bp in length
	148081	gap of unknown length
	148181	contig of 14550 bp in length
	148181	Location/Qualifiers
	source	
	1..162730	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="1"
		/clone="RP11-131015"
	misc_feature	
	1..2844	/note="assembly_name:Contig20"
	misc_feature	
	2945..4473	/note="assembly_name:Contig21"
	misc_feature	
	4574..6039	/note="assembly_name:Contig22"
	misc_feature	
	6140..7692	/note="assembly_name:Contig23"
	misc_feature	
	7793..10700	/note="assembly_name:Contig24"
	misc_feature	
	10801..13657	/note="assembly_name:Contig25"
	misc_feature	
	13758..17676	/note="assembly_name:Contig26"
	misc_feature	
	17777..22691	/note="assembly_name:Contig27"
	misc_feature	
	22792..26419	/note="assembly_name:Contig28"
	vector_side:left	
	26520..29391	/note="assembly_name:Contig29"
	misc_feature	
	29492..33271	/note="assembly_name:Contig30"
	misc_feature	
	33372..38256	/note="assembly_name:Contig31"
	misc_feature	
	38357..42491	/note="assembly_name:Contig32"
	misc_feature	
	42592..47107	/note="assembly_name:Contig33"
	misc_feature	
	47208..510	

```

/note="assembly_name:Contig39"
90782..10064
/note="assembly_name:Contig40"
100165..112323
/note="assembly_name:Contig41"
112424..122332
/note="assembly_name:Contig42"
122433..133727
/note="assembly_name:Contig43"
133828..148080
/note="assembly_name:Contig44"
148181..162730
/note="assembly_name:Contig45"
BASE COUNT 48389 a 32037 c 35685 g 47113 t 2506 others
ORIGIN

Query Match      2.1% Score 23; DB 51; Length 162730;
Best Local Similarity 100.0%; Pred. No. 0.16; Mismatches 0; Indels 0; Gaps 0;
Matches 23; Conservative 0;

QY 437 agtttgagactgctgggaaca 459
|||||
Db 64302 AGTTGAGACTGCGCTGGAACA 64280

RESULT 3
CNS06C7Z 172219 bp DNA HTG 07-AUG-2000
LOCUS Homo sapiens chromosome 14 clone C-2576L4, *** SEQUENCING IN
DEFINITION PROGRESS ***, 5 ordered pieces.
ACCESSION AL391153
VERSION AL391153.1 GI:9755851
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172219)
Genoscope.
Direct Submission
Submitted (07-AUG-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc. . even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.
Contigs composition :
20392 bp contig from 1 to 20392
22299 bp contig from 20493 to 42791
66708 bp contig from 42892 to 109599
6652 bp contig from 109700 to 116351
55768 bp contig from 116452 to 172219

Overall quality chart :
Range : bases
0 : 1403
1 - 9 : 105
10 - 19 : 587
20 - 29 : 878
30 - 39 : 1909
40 - 49 : 7802
50 - 59 : 11434
60 - 69 : 9192
70 - 79 : 14226
80 - 89 : 36725
90 - 99 : 87958

```

```

Percentage of bases with a quality value >= 40 : 97 %.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 20392: contig of 20392 bp in length
* 20393 20492: gap of 100 bp
* 20493 42791: contig of 22299 bp in length
* 42792 42891: gap of 100 bp
* 42892 109599: contig of 66708 bp in length
* 109600 109699: gap of 100 bp
* 109700 116351: contig of 6652 bp in length
* 116352 116451: gap of 100 bp
* 116452 172219: contig of 55768 bp in length.
FEATURES             Location/Qualifiers
     1..172219
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /chromosome="14"
     /clone_lib="CalTech-D"
     /clone="C-2576L4"
BASE COUNT 49858 a 36447 c 36087 g 49388 t 439 others
ORIGIN

Query Match      2.0% Score 22; DB 87; Length 172219;
Best Local Similarity 100.0%; Pred. No. 0.59; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;

QY 219 ttctccctccctctgtcca 240
|||||
Db 47356 TTTCTCCCTCCCTCTGTCCA 47377

RESULT 4
AC073443 61344 bp DNA HTG 18-JUN-2000
LOCUS Homo sapiens chromosome 3 clone RP11-147E24 map 3, LOW-PASS
DEFINITION SEQUENCE: SAMPLING.
ACCESSION AC073443
VERSION AC073443.1 GI:8571733
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 61344)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 3, clone RP11-147E24
Unpublished
2 (bases 1 to 61344)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McSheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,I., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

```

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testave,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (18-JUN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

-----  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www.seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 -----  
 Project Information  
 -----  
 Center project name: L10150  
 Center clone name: 147\_E\_24  
 -----

NOTE: This record contains 75 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 720 819: gap of 100 bp  
 820 1538: contig of 719 bp in length  
 1539 1638: gap of 100 bp  
 1639 2338: contig of 700 bp in length  
 2339 2438: gap of 100 bp  
 2439 3151: contig of 713 bp in length  
 3152 3251: gap of 100 bp  
 3252 3962: contig of 711 bp in length  
 3963 4062: gap of 100 bp  
 4063 4780: contig of 718 bp in length  
 4781 4880: gap of 100 bp  
 4881 5598: contig of 718 bp in length  
 5599 5698: gap of 100 bp  
 5699 6392: contig of 694 bp in length  
 6393 6492: gap of 100 bp  
 6493 7209: contig of 717 bp in length  
 7210 7309: gap of 100 bp  
 7310 8033: contig of 724 bp in length  
 8034 8133: gap of 100 bp  
 8134 8849: contig of 716 bp in length  
 8850 8949: gap of 100 bp  
 8950 9676: contig of 727 bp in length  
 9677 9776: gap of 100 bp  
 9777 10498: contig of 722 bp in length  
 10499 10598: gap of 100 bp  
 10599 11313: contig of 715 bp in length  
 11314 11413: gap of 100 bp  
 11414 12120: contig of 707 bp in length  
 12121 12220: gap of 100 bp  
 12221 12944: contig of 724 bp in length  
 12945 13044: gap of 100 bp  
 13045 13768: contig of 724 bp in length  
 13769 13868: gap of 100 bp  
 13869 14557: contig of 689 bp in length  
 14558 14657: gap of 100 bp  
 14658 15396: contig of 739 bp in length  
 15397 15496: gap of 100 bp  
 15497 16201: contig of 705 bp in length  
 16202 16301: gap of 100 bp  
 16302 17027: contig of 726 bp in length  
 17028 17127: gap of 100 bp  
 17128 17844: contig of 717 bp in length

17845 17944: gap of 100 bp  
 17945 18664: contig of 720 bp in length  
 18665 18764: gap of 100 bp  
 18765 19489: contig of 725 bp in length  
 19490 19589: gap of 100 bp  
 19590 20311: contig of 722 bp in length  
 20312 20411: gap of 100 bp  
 20412 21142: contig of 731 bp in length  
 21143 21242: gap of 100 bp  
 21243 21947: contig of 705 bp in length  
 21948 22047: gap of 100 bp  
 22048 22765: contig of 718 bp in length  
 22766 22865: gap of 100 bp  
 22866 23587: contig of 722 bp in length  
 23588 23687: gap of 100 bp  
 23688 24391: contig of 704 bp in length  
 24392 24491: gap of 100 bp  
 24492 25200: contig of 709 bp in length  
 25201 25300: gap of 100 bp  
 25301 26020: contig of 720 bp in length  
 26021 26120: gap of 100 bp  
 26121 26856: contig of 736 bp in length  
 26857 26956: gap of 100 bp  
 26957 27672: contig of 716 bp in length  
 27673 27772: gap of 100 bp  
 27773 28494: contig of 722 bp in length  
 28495 28594: gap of 100 bp  
 28595 29322: contig of 728 bp in length  
 29323 29422: gap of 100 bp  
 29423 30114: contig of 692 bp in length  
 30115 30214: gap of 100 bp  
 30215 30949: contig of 735 bp in length  
 30950 31049: gap of 100 bp  
 31050 31787: contig of 738 bp in length  
 31788 31887: gap of 100 bp  
 31888 32609: contig of 722 bp in length  
 32610 32709: gap of 100 bp  
 32710 33417: contig of 708 bp in length  
 33418 33517: gap of 100 bp  
 33518 34260: contig of 743 bp in length  
 34261 34360: gap of 100 bp  
 34361 35088: contig of 728 bp in length  
 35089 35188: gap of 100 bp  
 35189 35905: contig of 717 bp in length  
 35906 36005: gap of 100 bp  
 36006 36725: contig of 720 bp in length  
 36726 36825: gap of 100 bp  
 36826 37540: contig of 715 bp in length  
 37541 37640: gap of 100 bp  
 37641 38364: contig of 724 bp in length  
 38365 38464: gap of 100 bp  
 38465 39197: contig of 733 bp in length  
 39198 39297: gap of 100 bp  
 39298 40009: contig of 712 bp in length  
 40010 40109: gap of 100 bp  
 40110 40840: contig of 731 bp in length  
 40841 40940: gap of 100 bp  
 40941 41665: contig of 725 bp in length  
 41666 41765: gap of 100 bp  
 41766 42497: contig of 732 bp in length  
 42498 42597: gap of 100 bp  
 42598 43320: contig of 723 bp in length  
 43321 43420: gap of 100 bp  
 43421 44140: contig of 720 bp in length  
 44141 44240: gap of 100 bp  
 44241 44948: contig of 708 bp in length  
 44949 45048: gap of 100 bp  
 45049 45785: contig of 737 bp in length  
 45786 45885: gap of 100 bp  
 45886 46587: contig of 702 bp in length  
 46588 46687: gap of 100 bp  
 46688 47406: contig of 719 bp in length  
 47407 47506: gap of 100 bp

```

* 47507 48220: contig of 714 bp in length
* 48221 48320: gap of 100 bp
* 48321 49049: contig of 729 bp in length
* 49050 49149: gap of 100 bp
* 49150 49863: contig of 714 bp in length
* 49864 49963: gap of 100 bp
* 49964 50687: contig of 724 bp in length
* 50688 50787: gap of 100 bp
* 50788 51513: contig of 726 bp in length
* 51514 51613: gap of 100 bp
* 51614 52348: contig of 735 bp in length
* 52349 52448: gap of 100 bp
* 52449 53173: contig of 725 bp in length
* 53174 53273: gap of 100 bp
* 53274 54001: contig of 728 bp in length
* 54002 54101: gap of 100 bp
* 54102 54828: contig of 727 bp in length
* 54829 54928: gap of 100 bp
* 54929 55628: contig of 700 bp in length
* 55629 55728: gap of 100 bp
* 55729 56450: contig of 722 bp in length
* 56451 56550: gap of 100 bp

Query Match 1.9%; Score 21; DB 57; Length 61344;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 agttgagactggcctgggaa 457
|||||
Db 23182 AGTTGAGACTGGCTGGGAA 23202

RESULT 5
AC005522
LOCUS
DEFINITION Homo sapiens clone RP5-1129E22, complete sequence.
ACCESSION AC005522
VERSION AC005522.2 GI:7770659
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149605)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 149605)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 149605)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 12, 2000 this sequence version replaced gi:3907491.
Center project name: H_DJ1129E22.
FEATURES
source
1. 149605
/organism="Homo sapiens"
/dbxref="taxon:9606"
/clone="RP5-1129E22"
BASE COUNT 37830 a 38344 c 37681 g 35750 t

```

## ORIGIN

```

Query Match 1.9%; Score 21; DB 9; Length 149605;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 agttgagactggcctgggaa 457
|||||
Db 71853 AGTTGAGACTGGCTGGGAA 71873

RESULT 6
AC023408
LOCUS
DEFINITION Homo sapiens chromosome 6 clone RP11-612M16 map 6, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
ACCESSION AC023408
VERSION AC023408.2 GI:8316895
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 153422)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 6, clone RP11-612M16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 153422)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepl,i., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,I.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivat,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 7, 2000 this sequence version replaced gi:6970573.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6363
Center clone name: 612_M_16
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 135238 bases at least Q40
Consensus quality: 141969 bases at least Q30
Consensus quality: 146244 bases at least Q20
Insert size: 150000; agarose-fp

```

Insert size: 151322; sum-of-contigs  
Quality coverage: 4.0 in Q20 bases; agarose-gp  
Quality coverage: 3.9 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 1153 1252: gap of 100 bp  
\* 1253 2252: contig of 1152 bp in length  
\* 2253 2352: gap of 100 bp  
\* 2353 3503: contig of 1000 bp in length  
\* 3504 3603: gap of 100 bp  
\* 3604 4658: contig of 1151 bp in length  
\* 4659 4758: gap of 100 bp  
\* 4759 6059: contig of 1055 bp in length  
\* 6060 6159: gap of 100 bp  
\* 6160 6832: contig of 1301 bp in length  
\* 6833 6932: gap of 100 bp  
\* 6933 9547: contig of 673 bp in length  
\* 9548 9647: gap of 100 bp  
\* 9648 11267: contig of 2615 bp in length  
\* 11268 11367: gap of 100 bp  
\* 11368 14730: contig of 1620 bp in length  
\* 14731 14830: gap of 100 bp  
\* 14831 18501: contig of 3363 bp in length  
\* 18502 18601: gap of 100 bp  
\* 18602 23876: contig of 3671 bp in length  
\* 23877 23976: gap of 100 bp  
\* 23977 28125: contig of 5275 bp in length  
\* 28126 28225: gap of 100 bp  
\* 28226 33384: contig of 4149 bp in length  
\* 33385 33484: gap of 100 bp  
\* 33485 37950: contig of 5159 bp in length  
\* 37951 38050: gap of 100 bp  
\* 38051 43278: contig of 4466 bp in length  
\* 43279 43378: gap of 100 bp  
\* 43379 45641: contig of 5228 bp in length  
\* 45642 49741: gap of 100 bp  
\* 49742 57239: contig of 6263 bp in length  
\* 57240 57339: gap of 100 bp  
\* 57340 65126: contig of 7498 bp in length  
\* 65127 65226: gap of 100 bp  
\* 65227 74642: contig of 7787 bp in length  
\* 74643 74742: gap of 100 bp  
\* 74743 86960: contig of 9416 bp in length  
\* 86961 87060: gap of 100 bp  
\* 87061 115886: contig of 12218 bp in length  
\* 115887 115986: gap of 100 bp  
\* 115987 153422: contig of 28826 bp in length  
\* 153423 37436: contig of 37436 bp in length.

FEATURES  
Source  
1. 153422  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6"  
/clone="RP11-612M16"  
1. 1152  
/clone\_lib="RC1-11 Human Male BAC"  
misc\_feature  
1253. 2252  
/note="assembly\_fragment"  
misc\_feature  
2353. 3503  
/note="assembly\_fragment"  
misc\_feature  
3604. 4658  
/note="assembly\_fragment"  
misc\_feature  
4759. 6059  
/note="assembly\_fragment"

misc\_feature  
6160. 6832  
/note="assembly\_fragment"  
clone\_end:17  
vector\_side:right  
6933. 9547  
/note="assembly\_fragment"  
9648. 11267  
/note="assembly\_fragment"  
11368. 14730  
/note="assembly\_fragment"  
14831. 18501  
/note="assembly\_fragment"  
18602. 23876  
/note="assembly\_fragment"  
23977. 28125  
/note="assembly\_fragment"  
28226. 33384  
/note="assembly\_fragment"  
33485. 37950  
/note="assembly\_fragment"  
38051. 43278  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
43379. 49641  
/note="assembly\_fragment"  
49742. 57239  
/note="assembly\_fragment"  
57340. 65126  
/note="assembly\_fragment"  
65227. 74642  
/note="assembly\_fragment"  
74743. 86960  
/note="assembly\_fragment"  
87061. 115886  
/note="assembly\_fragment"  
115987. 153422  
/note="assembly\_fragment"  
BASE COUNT 47446 a 29588 c 28599 g 45687 t 2102 others  
ORIGIN

Query Match 1.9%; Score 21; DB 49; Length 153422;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 437 agtttgagactggcctggaa 457  
DB 76705 AGTTTGAGACTGGCCTGGAA 76725

RESULT 7  
AC023851/c  
LOCUS AC023851.2 155655 bp DNA HTG 08-MAR-2000  
DEFINITION Homo sapiens clone RP11-6417, WORKING DRAFT SEQUENCE, 10 unordered  
pieces.  
ACCESSION AC023851  
VERSION AC023851.2 GI:7209989  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 155655)  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Homo sapiens, clone RP11-6417  
JOURNAL Unpublished  
2 (bases 1 to 155655)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bada, F., Boguslavsky, L.,  
Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,  
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
Dearrellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,

Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Govette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L., Kiehl, J.C., Iliiev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, S., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (18-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 8, 2000 this sequence version replaced gi:6997299.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
Center project name: L6123

Center clone name: 64\_I-7

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 144832 bases at least Q40

Consensus quality: 151535 bases at least Q30

Consensus quality: 153767 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 154755; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; agarose-fp

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2405: contig of 2405 bp in length  
\* 2406 2505: gap of 100 bp  
\* 2506 3884: contig of 1379 bp in length  
\* 3885 3984: gap of 100 bp  
\* 3985 6557: contig of 2573 bp in length  
\* 6558 6657: gap of 100 bp  
\* 6658 10024: contig of 3367 bp in length  
\* 10025 10124: gap of 100 bp  
\* 10125 23105: contig of 12981 bp in length  
\* 23106 23205: gap of 100 bp  
\* 23206 36867: contig of 13662 bp in length  
\* 36868 36967: gap of 100 bp  
\* 36968 54767: contig of 17799 bp in length  
\* 54767 54866: gap of 100 bp  
\* 54867 79648: contig of 24782 bp in length  
\* 79649 79748: gap of 100 bp  
\* 79749 118400: contig of 38652 bp in length  
\* 118401 118500: gap of 100 bp  
\* 118501 155655: contig of 37155 bp in length.  
\* Location/Qualifiers  
\* 1..155655  
\* /organism="Homo sapiens"

FEATURES  
SOURCE

/db\_xref="taxon:9606"  
/clone="RP11-6417"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..2405  
/note="assembly\_fragment"  
2506..3884  
/note="assembly\_fragment"  
3985..6557  
/note="assembly\_fragment"  
6658..10024  
/note="assembly\_fragment"  
10125..23105  
/note="assembly\_fragment"  
23206..36867  
/note="assembly\_fragment"  
36968..54766  
/note="assembly\_fragment"  
54867..79648  
/note="assembly\_fragment"  
79749..118400  
/note="assembly\_fragment"  
clone\_end:77  
vector\_side:right  
118501..155655  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
47723 a 30373 c 30013 g 46645 t 901 others  
BASE COUNT  
ORIGIN

Query Match 1.9%; Score 21; DB 49; Length 155655;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 ttctccctccctccctgtctg 239  
|||||

Db 112589 TTCTCCCTCCCTCCCTGTCC 112569

RESULT 8  
AL355812/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL355812 155849 bp DNA HTG 22-MAY-2000  
Homo sapiens chromosome X clone RP11-81003, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*; 19 unordered pieces.

AL355812 AL355812

AL355812.3 GI:9213648

HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 155849)

Pavitt, R.

Direct Submission

Submitted (22-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)

On Jul 15, 2000 this sequence version replaced gi:8052114.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

----- Project Information

Center project name: BA81003

----- Summary Statistics

Sequencing program: XGAP4; version 4.5

Chemistry: Dye-terminator ET-amersham; 7% of reads

Dye-terminator Big Dye; 92% of reads

Consensus quality: 145075 bases at least Q40

Consensus quality: 149049 bases at least Q30

## us-09-101-423a-4.olig.rge

Mon Nov 6 10:14:24 2000

Consensus quality: 151509 bases at least Q20  
 Insert size: 154049; sum-of-contigs  
 Insert size: 138402; agarose-fp  
 Quality coverage: 3.36x in Q20 bases; sum-of-contigs Quality  
 coverage: 3.74x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
 consists of 19 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1574: contig of 1574 bp in length  
 1575 1674: gap of 100 bp  
 1675 6811: contig of 5137 bp in length  
 6812 6911: gap of 100 bp  
 6912 8434: contig of 1523 bp in length  
 8435 8534: gap of 100 bp  
 8535 22976: contig of 14442 bp in length  
 22977 23076: gap of 100 bp  
 23077 28229: contig of 5153 bp in length  
 28230 28329: gap of 100 bp  
 28330 30623: contig of 2294 bp in length  
 30624 30723: gap of 100 bp  
 30724 37857: contig of 7134 bp in length  
 37858 37957: gap of 100 bp  
 37958 34758: contig of 16801 bp in length  
 34759 54858: gap of 100 bp  
 54859 66398: contig of 11540 bp in length  
 66399 66498: gap of 100 bp  
 66499 76109: contig of 9611 bp in length  
 76110 76209: gap of 100 bp  
 76210 80280: contig of 4071 bp in length  
 80281 80380: gap of 100 bp  
 80381 90473: contig of 10093 bp in length  
 90474 90573: gap of 100 bp  
 90574 110756: contig of 20183 bp in length  
 110757 110856: gap of 100 bp  
 110857 117605: contig of 6749 bp in length  
 117606 117705: gap of 100 bp  
 117706 121925: contig of 4220 bp in length  
 121926 122026: gap of 100 bp  
 122026 128877: contig of 6852 bp in length  
 128878 128977: gap of 100 bp  
 128978 141150: contig of 12173 bp in length  
 141151 141250: gap of 100 bp  
 141251 154381: contig of 13131 bp in length  
 154382 154481: gap of 100 bp  
 154482 155849: contig of 1368 bp in length.

## FEATURES

source  
 1..155849  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /clone="RP11-81003"  
 /clone\_lib="RPC1-11.3"  
 1..1574  
 /note="assembly\_fragment:00296  
 fragment\_chain:1"  
 1675..6811  
 /note="assembly\_fragment:00780  
 fragment\_chain:1"  
 6912..8434  
 /note="assembly\_fragment:00746  
 fragment\_chain:1"  
 8535..22976  
 /note="assembly\_fragment:00583  
 fragment\_chain:1"  
 23077..28229  
 /note="assembly\_fragment:00047  
 fragment\_chain:1"

misc\_feature  
 28330..30623  
 /note="assembly\_fragment:01138  
 fragment\_chain:1"  
 misc\_feature  
 30724..37857  
 /note="assembly\_fragment:01098  
 fragment\_chain:1"  
 misc\_feature  
 37958..54758  
 /note="assembly\_fragment:00155  
 fragment\_chain:1"  
 misc\_feature  
 54859..66398  
 /note="assembly\_fragment:01205  
 fragment\_chain:1"  
 misc\_feature  
 66499..76109  
 /note="assembly\_fragment:00423  
 fragment\_chain:2"  
 misc\_feature  
 76210..80280  
 /note="assembly\_fragment:00515  
 fragment\_chain:2"  
 misc\_feature  
 80381..90473  
 /note="assembly\_fragment:00428  
 fragment\_chain:3"  
 misc\_feature  
 90574..110756  
 /note="assembly\_fragment:01012  
 fragment\_chain:3"  
 misc\_feature  
 110857..117605  
 /note="assembly\_fragment:00918  
 fragment\_chain:4"  
 misc\_feature  
 117706..121925  
 /note="assembly\_fragment:01072  
 fragment\_chain:4"  
 misc\_feature  
 122026..128877  
 /note="assembly\_fragment:00586"  
 fragment\_chain:5"  
 misc\_feature  
 128978..141150  
 /note="assembly\_fragment:00852"  
 fragment\_chain:5"  
 misc\_feature  
 141251..154381  
 /note="assembly\_fragment:00545  
 fragment\_chain:5"  
 misc\_feature  
 154482..155849  
 /note="assembly\_fragment:00824  
 fragment\_chain:5"  
 clone\_end:T7  
 vector\_side:right  
 46750 a 28742 c 27854 g 50672 t 1831 others

## BASE COUNT

Query Match 1.9%; Score 21; DB 84; Length 155849;  
 Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 686 atgagttcatggctttttttt 706

|||||  
 Db 10052 ATGAGTTCAIGGCTTTTATTT 10032

## RESULT

9  
 AC023369  
 LOCUS  
 DEFINITION  
 AC023369  
 VERSION  
 AC023369.3 GI:8072542  
 KEYWORDS  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 163300)  
 Birren,E., Linton,L., Nusbaum,C. and Lander,E.  
 Unpublished  
 REFERENCE  
 2 (bases 1 to 163300)  
 Birren,E., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

## AC023369

HTG  
 26-MAY-2000

## DEFINITION

AC023369

## VERSION

AC023369.3

## KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

## SOURCE

human.

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 163300)

## AUTHORS

Birren,E., Linton,L., Nusbaum,C. and Lander,E.

## TITLE

Unpublished

## JOURNAL

REFERENCE

2 (bases 1 to 163300)

Birren,E., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

TITLE	JOURNAL	COMMENT
-------	---------	---------

```

Center clone name: 378_A10
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 144227 bases at least Q40
Consensus quality: 153621 bases at least Q30
Consensus quality: 157291 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 160400; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.7 in Q20 bases; sum-of-contigs

```

	1	1074:	contig	of 1074	bp in length
1075	1174:	gap	of	100	bp
1175	2825:	contig	of 1651	bp in length	
2826	2925:	gap	of	100	bp
2926	5468:	contig	of 2543	bp in length	
5469	5568:	gap	of	100	bp
5569	7045:	contig	of 1477	bp in length	
7046	7145:	gap	of	100	bp
7146	8995:	contig	of 2750	bp in length	
8996	9995:	gap	of	100	bp
9996	13441:	contig	of 3448	bp in length	
13442	13541:	gap	of	100	bp
13542	15891:	contig	of 2350	bp in length	
15892	15991:	gap	of	100	bp
15992	1854:	contig	of 2556	bp in length	
1854	18647:	gap	of	100	bp
18648	21908:	contig	of 3261	bp in length	
21909	22008:	gap	of	100	bp



```

misc_feature 30188..33771
/note="assembly_fragment"
misc_feature 33872..37601
/note="assembly_fragment"
misc_feature 37702..41824
/note="assembly_fragment"
misc_feature 41925..45491
/note="assembly_fragment"
misc_feature 45592..50357
/note="assembly_fragment"
misc_feature 50458..56039
/note="assembly_fragment"
misc_feature 56140..61103
/note="assembly_fragment"
misc_feature 61204..66481
/note="assembly_fragment"
misc_feature 66582..72051
/note="assembly_fragment"
misc_feature 72152..79292
/note="assembly_fragment"
misc_feature 79393..85677
/note="assembly_fragment"
misc_feature 85778..94322
/note="assembly_fragment"
misc_feature 94423..103710
/note="assembly_fragment"
misc_feature 103811..109535
/note="assembly_fragment"
clone_end:T7
vector_side:right
misc_feature 109636..119773
/note="assembly_fragment"
misc_feature 119874..131960
/note="assembly_fragment"
misc_feature 132061..147678
/note="assembly_fragment"
clone_end:SP6
vector_side:left

Query Match 1.9%; Score 21; DB 49; Length 163300;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 agtttgagactgctgggaa 457
Db 107880 AGTTTGAGACTGCTGGGAA 107900

RESULT 10
AC026830 166484 bp DNA HTG 19-APR-2000
LOCUS Homo sapiens chromosome 1 clone RP11-215H14 map 1, WORKING DRAFT
SEQUENCE, 32 unordered pieces.
ACCESSION AC026830
VERSION AC026830.2 GI:7596890
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166484)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckghalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

```

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, F., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trifilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (24-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 19, 2000 this sequence version replaced 91.7321633.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L7706  
Center clone name: 215\_H-14  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 150453 bases at least Q40  
Consensus quality: 158028 bases at least Q30  
Consensus quality: 161192 bases at least Q20  
Insert size: 176000; agarose-IP  
Insert size: 163384; sum-of-contigs  
Quality coverage: 3.6 in Q20 bases; agarose-IP  
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1024: contig of 1024 bp in length  
1025 1124: gap of 100 bp  
1125 2485: contig of 1361 bp in length  
2486 2585: gap of 100 bp  
2586 3626: contig of 1041 bp in length  
3627 3726: gap of 100 bp  
3727 4884: contig of 1158 bp in length  
4885 4984: gap of 100 bp  
4985 6513: contig of 1529 bp in length  
6514 6613: gap of 100 bp  
6614 8513: contig of 1900 bp in length  
8514 8613: gap of 100 bp  
8614 10616: contig of 2003 bp in length  
10617 10716: gap of 100 bp  
10717 12445: contig of 1729 bp in length  
12446 12545: gap of 100 bp  
12546 16043: contig of 3498 bp in length  
16044 16143: gap of 100 bp  
16144 19800: contig of 3657 bp in length  
19801 19900: gap of 100 bp  
19901 23755: contig of 3855 bp in length  
23756 23855: gap of 100 bp  
23856 25033: contig of 1178 bp in length  
25034 25133: gap of 100 bp  
25134 28296: contig of 3163 bp in length

```
* 28297 28396: gap of 100 bp
* 28397 31054: contig of 2658 bp in length
* 31055 31154: gap of 100 bp
* 31155 35420: contig of 4266 bp in length
* 35421 35520: gap of 100 bp
* 35521 39220: contig of 3700 bp in length
* 39221 39320: gap of 100 bp
* 39321 42676: contig of 3356 bp in length
* 42677 42776: gap of 100 bp
* 42777 46418: contig of 3642 bp in length
* 46419 46518: gap of 100 bp
* 46519 51021: contig of 4503 bp in length
* 51022 51121: gap of 100 bp
* 51122 56073: contig of 4952 bp in length
* 56074 56173: gap of 100 bp
* 56174 61048: contig of 4875 bp in length
* 61049 61148: gap of 100 bp
* 61149 66205: contig of 5057 bp in length
* 66206 66306: gap of 100 bp
* 66306 72086: contig of 5781 bp in length
* 72087 72186: gap of 100 bp
* 72187 77190: contig of 5004 bp in length
* 77191 77290: gap of 100 bp
* 77291 87414: contig of 10124 bp in length
* 87415 87514: gap of 100 bp
* 87515 94355: contig of 6841 bp in length
* 94356 94455: gap of 100 bp
* 94456 103495: contig of 9040 bp in length
* 103496 103595: gap of 100 bp
* 103596 112676: contig of 9081 bp in length
* 112677 112776: gap of 100 bp
* 112777 122142: contig of 9366 bp in length
* 122143 122242: gap of 100 bp
* 122243 131715: contig of 9473 bp in length
* 131716 131815: gap of 100 bp
* 131816 142646: contig of 10831 bp in length
* 142647 142746: gap of 100 bp
* 142747 166484: contig of 23738 bp in length.
```

## FEATURES

source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1"
/clone="RP11-215H14"
/clone_lib="RPC1-11 Human Male BAC"
1. .1024
/note="assembly_fragment"
1125. 2485
/note="assembly_fragment"
2586. 3626
/note="assembly_fragment"
3727. 4884
/note="assembly_fragment"
4985. 6513
/note="assembly_fragment"
6614. 8513
/note="assembly_fragment"
8614. 10616
/note="assembly_fragment"
10717. 12445
/note="assembly_fragment"
12546. 16043
/note="assembly_fragment"
16144. 19800
/note="assembly_fragment"
19901. 23755
/note="assembly_fragment"
23856. 25033
/note="assembly_fragment"
clone_end:17
vector_side:left
25134. 28296
```

```
misc_feature
28397. 31054
/note="assembly_fragment"
misc_feature
31155. 35420
/note="assembly_fragment"
misc_feature
35521. 39220
/note="assembly_fragment"
misc_feature
39321. 42676
/note="assembly_fragment"
misc_feature
42777. 46418
/note="assembly_fragment"
misc_feature
46519. 51021
/note="assembly_fragment"
misc_feature
51122. 56073
/note="assembly_fragment"
misc_feature
56174. 61048
/note="assembly_fragment"
misc_feature
61149. 66205
/note="assembly_fragment"
clone_end:SP6
vector_side:right
66306. 72086
/note="assembly_fragment"
misc_feature
72187. 77190
/note="assembly_fragment"
misc_feature
77291. 87414
/note="assembly_fragment"
misc_feature
87515. 94355
/note="assembly_fragment"
misc_feature
94456. 103495
/note="assembly_fragment"
misc_feature
103496. 112676
/note="assembly_fragment"
```

Query Match 1.9%; Score 21; DB 51; Length 166484;

Matches Local Similarity 100.0%; Pred.No.2.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 agtttgagactggcctgggaa 457

|||||

Db 4555 ACTTTGAGACIGCCTGGGAA 4575

RESULT 11

AL162253/C

LOCUS

DEFINITION Homo sapiens chromosome 9 clone RP11-574F11, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 11 unordered pieces.

ACCESSION AL162253

VERSION AL162253.8 GI:9369080

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Plumb,B.

Direct Submission

Submitted (19-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jul 22, 2000 this sequence version replaced gi:9368689.

COMMENT

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA574F11

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 164990 bases at least Q40

Consensus quality: 167044 bases at least Q30

Consensus quality: 168529 bases at least Q20  
Insert size: 170319; sum-of-contigs  
Quality coverage: 9.5% error; agarose-fp  
Quality coverage: 5.99% in Q20 bases; sum-of-contigs Quality  
coverage: 6.14% in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 79724: contig of 79724 bp in length  
\* 79725 79824: gap of 100 bp  
\* 79825 88916: contig of 9092 bp in length  
\* 88917 89016: gap of 100 bp  
\* 89017 97692: contig of 8676 bp in length  
\* 97693 97792: gap of 100 bp  
\* 97793 101702: contig of 3910 bp in length  
\* 101703 101802: gap of 100 bp  
\* 101803 103207: contig of 1405 bp in length  
\* 103208 103307: gap of 100 bp  
\* 103308 104396: contig of 1089 bp in length  
\* 104397 104496: gap of 100 bp  
\* 104497 105723: contig of 1227 bp in length  
\* 105724 105823: gap of 100 bp  
\* 105824 107112: contig of 1289 bp in length  
\* 107113 107212: gap of 100 bp  
\* 107213 108535: contig of 1323 bp in length  
\* 108536 108635: gap of 100 bp  
\* 108636 170204: contig of 61569 bp in length  
\* 170205 170304: gap of 100 bp  
\* 170305 171319: contig of 1015 bp in length.

Location/Qualifiers  
1. 171319  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-574F11"  
/clone\_lib="RPC1-11.2"  
1. 79724  
/note="assembly\_fragment:00479  
fragment\_chain:1"  
79825..88916  
/note="assembly\_fragment:01837  
fragment\_chain:1"  
89017..97692  
/note="assembly\_fragment:02014  
fragment\_chain:2"  
97793..101702  
/note="assembly\_fragment:01533  
fragment\_chain:2"  
101803..103207  
/note="assembly\_fragment:00488"  
103308..104396  
/note="assembly\_fragment:00522"  
104497..105723  
/note="assembly\_fragment:00911"  
105824..107112  
/note="assembly\_fragment:01210"  
107213..108535  
/note="assembly\_fragment:01565"  
108636..170204  
/note="assembly\_fragment:01644"  
170305..171319  
/note="assembly\_fragment:02077"  
51583 a 34397 c 34512 g 49827 t 1000 others

BASE COUNT 51583 a 34397 c 34512 g 49827 t 1000 others  
ORIGIN  
Query Match 1.9%; Score 21; DB 64; Length 171319;

Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 437 agtttgagactggcctgggaa 457  
|||||  
Db 91356 AGTTTGAGACTGGCCTGGGAA 91336

RESULT 12  
AC004021/c  
LOCUS AC004021 171370 bp DNA PRI 04-FEB-2000  
DEFINITION Human PAC clone RP1-186K10 from 5q31, complete sequence.  
ACCESSION AC004021  
VERSION AC004021.1 GI:2795824  
KEYWORDS HTG.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 171370)  
AUTHORS Kalicki, J., Kellen, J. and O'Brien, D.  
TITLE The sequence of H. sapiens PAC clone RP1-186K10  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 171370)  
AUTHORS Waterston, R.  
TITLE Direct Submission

JOURNAL Submitted (20-JAN-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 3 (bases 1 to 171370)  
AUTHORS Waterston, R.  
TITLE Direct Submission

JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE 4 (bases 1 to 171370)  
AUTHORS Waterston, R.  
TITLE Direct Submission

JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics -----  
Center project name: H\_DJ0186K10

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone; and the assembly was  
confirmed by restriction digest.

MAPPING INFORMATION:  
This clone was provided for sequencing by Dr. Michelle M. Le Beau  
in the Department of Medicine, University of Chicago, Chicago IL;  
and Dr. John D. McPherson, Department of Genetics, Washington  
University, St. Louis MO. For additional information about the map  
position of this clone, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
This clone was derived from human PAC library RPC1-1, prepared by  
Pieter de Jong and coworkers at Roswell Park Cancer Institute,  
using the method described by Ioannou et al., Nature Genetics,  
6:84-9 (1994). The library is from one male donor. For further  
details, see <http://bacpac.med.buffalo.edu/>  
The clone is available from Genome Systems, Inc.

(http://www.genomesystems.com).

VECTOR: pCYRAC2

#### NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RPI-186K10; actual end is at 171370 of RPI-186K10. The orientation of this clone is unknown.

This clone contains STS AFM350ybl (NID:g1051891) and EST53547 (NID:g1347444).

#### FEATURES

```
source      Location/Qualifiers
1..171370
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="5"
  /clone_lib="RPC1-1"
  /map="5q31"
  1..1157
    /rpt_family="L1"
    1290..2612
      /rpt_family="L1"
      2617..2679
        /rpt_family="MER81"
        3511..3643
          /rpt_family="MER1_type"
          3791..4239
            /rpt_family="L2"
            4245..4551
              /rpt_family="MaLR"
              4553..4936
                /rpt_family="Retroviral"
                5017..5328
                  /rpt_family="MER2_type"
                  5361..5408
                    /rpt_family="Retroviral"
                    5409..5542
                      /rpt_family="MaLR"
                      5546..5893
                        /rpt_family="L2"
                        5920..5994
                          /rpt_family="MaLR"
                          6030..6172
                            /rpt_family="L2"
                            6193..6256
                              /rpt_family="L2"
                              6299..6504
                                /rpt_family="L2"
                                6505..6536
                                  /rpt_family="(TGAA)n"
                                  7620..8444
                                    /rpt_family="L1"
                                    8919..8998
                                      /rpt_family="Retroviral"
                                      9041..10574
                                        /note="CpG island (GC=73.7, o/e=0.70, #CpGs=115)"
                                        10335..10420
                                          /rpt_family="Alu"
                                          10416..10523
                                            /rpt_family="Alu"
                                            10535..10594
                                              /rpt_family="Alu"
                                              10732..10988
                                                /rpt_family="L1"
                                                11023..14892
                                                  /rpt_family="L1"
                                                  1585..15760
                                                    /rpt_family="MER1_type"
                                                    18900..18902
                                                      /rpt_family="MIR"
                                                      18903..18949
                                                        /rpt_family="L2"
                                                        18950..18972
                                                          /rpt_family="MIR"
```

```
repeat_region 19618..19768 /rpt_family="L1"
repeat_region 19769..20895 /rpt_family="L1"
repeat_region 20863..22033 /rpt_family="L1"
repeat_region 22147..24390 /rpt_family="L1"
repeat_region 24614..25397 /rpt_family="L1"
repeat_region 25396..26577 /rpt_family="L1"
repeat_region 27204..27824 /rpt_family="L1"
repeat_region 28298..28456 /rpt_family="Alu"
repeat_region 28460..28510 /rpt_family="7SLRNA"
repeat_region 28548..28834 /rpt_family="Alu"
repeat_region 28855..29167 /rpt_family="Alu"
repeat_region 29768..30055 /rpt_family="L1"
repeat_region 31184..31477 /rpt_family="Alu"
repeat_region 32124..32489 /rpt_family="MaLR"
repeat_region 32647..32733 /rpt_family="MIR"
repeat_region 33532..33594 /rpt_family="L2"
repeat_region 33613..33706 /rpt_family="L1"
repeat_region 34315..34609 /rpt_family="Alu"
repeat_region 34782..34932 /rpt_family="L2"
repeat_region 34935..35152 /rpt_family="MIR"
repeat_region 36512..36802 /rpt_family="Alu"
repeat_region 36882..37126 /rpt_family="L2"
repeat_region 37957..38046 /rpt_family="L2"
repeat_region 38742..38863 /rpt_family="L2"
repeat_region 40308..40599 /rpt_family="MIR"
repeat_region 41127..41200 /rpt_family="Alu"
repeat_region 41356..41728 /rpt_family="MaLR"
repeat_region 41917..42152 /rpt_family="L2"
repeat_region 42169..42459 /rpt_family="Alu"
repeat_region 42468..42623 /rpt_family="L2"
repeat_region 43252..43548 /rpt_family="Alu"
repeat_region 43535..43594 /rpt_family="(TAA)n"
```

Query Match 1.9%; Score 21; DB 9; Length 171370;  
Best Local Similarity 100.0%; Pred.No.2.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 ggactgctaaattggggaggca 95

|||||  
DB 54830 GGACTGCTAAATTGGGGAGGCA 54810

## RESULT 13

AC021623/c

LOCUS

DEFINITION

AC021623 173175 bp DNA HTG 12-MAR-2000  
Homo sapiens clone RP11-365P20, WORKING DRAFT SEQUENCE, 12  
unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

AC021623.3 GI:7230215  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 173175)  
Biren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens, clone RP11-365P20

2 (bases 1 to 173175)

Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Becker, R., Beda, F.,  
Boguslavsky, L., Boukhgaiter, B., Brown, A., Burkett, G., Castle, A.,  
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenesstor, J.,  
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheeters, R., Meldrim, J., Menues, L., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,  
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,  
Zimmer, A. and Zody, M.

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 12, 2000 this sequence version replaced GI:6899690.

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5872

Center clone name: 365\_P\_20

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 166039 bases at least Q40

Consensus quality: 169053 bases at least Q30

Consensus quality: 170316 bases at least Q20

Insert size: 170000; agarose-gel

Insert size: 172075; sum-of-contigs

Quality coverage: 5.3 in Q20 bases; agarose-fp

Quality coverage: 5.3 in Q20 bases; sum-of-contigs

-----

NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 363: contig of 363 bp in length

\* 364 463: gap of 100 bp

\* 464 3209: contig of 2746 bp in length

\* 3210 3309: gap of 100 bp  
\* 3310 6561: contig of 3252 bp in length  
\* 6562 6661: gap of 100 bp  
\* 6662 11626: contig of 4965 bp in length  
\* 11627 11726: gap of 100 bp  
\* 11727 18147: contig of 6421 bp in length  
\* 18148 18247: gap of 100 bp  
\* 18248 26581: contig of 8334 bp in length  
\* 26582 26681: gap of 100 bp  
\* 26682 36588: contig of 9907 bp in length  
\* 36589 36688: gap of 100 bp  
\* 36689 48911: contig of 12223 bp in length  
\* 48912 49011: gap of 100 bp  
\* 49012 66671: contig of 17660 bp in length  
\* 66672 66771: gap of 100 bp  
\* 66772 91485: contig of 24714 bp in length  
\* 91486 91585: gap of 100 bp  
\* 91586 124630: contig of 33045 bp in length  
\* 124631 124730: gap of 100 bp  
\* 124731 173175: contig of 48445 bp in length.

## FEATURES

source

1..173175  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-365P20"  
/clone\_lib="RPCI-11 Human Male BAC"

misc\_feature

1..363  
/note="assembly\_fragment"  
vector\_end:17  
vector\_side:right

misc\_feature

464..3209  
/note="assembly\_fragment"  
3310..6561  
/note="assembly\_fragment"

misc\_feature

6662..11626  
/note="assembly\_fragment"  
11727..18147  
/note="assembly\_fragment"

misc\_feature

18248..26581  
/note="assembly\_fragment"  
26682..36588  
/note="assembly\_fragment"

misc\_feature

36689..48911  
/note="assembly\_fragment"  
49012..66671  
/note="assembly\_fragment"

misc\_feature

66772..91485  
/note="assembly\_fragment"  
91586..124630  
/note="assembly\_fragment"

misc\_feature

124731..173175  
vector\_side:right  
/note="assembly\_fragment"

misc\_feature

55147 a 29644 c 30543 g 56738 t  
1103 others

ORIGIN

Query Match

Best Local Similarity 1.9%; Score 21; DB 47; Length 173175;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 agttgagactggcctggaa 457

|||||

Db 59303 AGTTTGAGACGGCTGGGAA 59283

RESULT 14

AC068712/c

LOCUS

DEFINITION

DRAFT SEQUENCE, 22 unordered pieces.

ACCESSION

AC068712 175340 bp DNA HTG 03-JUL-2000  
Homo sapiens chromosome 15 clone CTD-3203H14 map 15q21.3, WORKING  
DRAFT SEQUENCE, 22 unordered pieces.

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC068712.4 GI:8901196  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
human.

REFERENCE  
AUTHORS

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 175340)  
Roven, L., Madan, A., Qin, S., Abbasi, N., Baradarani, L., Birditt, B.,  
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A.,  
Madan, A., Nesbitt, R., Shafer, T. and Hood, L.  
Sequencing of human chromosome 15 D15S146-D15S117 region  
Unpublished

TITLE  
JOURNAL  
AUTHORS

2 (bases 1 to 175340)  
Roven, L., Madan, A., Qin, S., Abbasi, N., Baradarani, L., Birditt, B.,  
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A.,  
Madan, A., Nesbitt, R., Shafer, T. and Hood, L.  
Direct Submission

TITLE  
JOURNAL

Submitted (07-MAY-2000) Multimegabase Sequencing Center, Institute  
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA  
98105, USA

COMMENT

On Jul 3, 2000 this sequence version replaced gi:8570425.  
----- Genome Center

Center: Multimegabase Sequencing Center

Center code: DNMSC

Web site: [http://chroma.mbt.washington.edu/msg\\_www](http://chroma.mbt.washington.edu/msg_www)

Contact: [leerowensystemsbiology.org](mailto:leerowensystemsbiology.org)

----- Summary Statistics

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-primer Big Dye; 10% of reads

Assembly program: Phrap; version 0.990399

Insert size: 180000; agarose-fp

Quality coverage: 4.0x in Q20 bases; sum-of-contigs

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1661: contig of 1661 bp in length  
1761: gap of unknown length  
1762 4466: contig of 2705 bp in length  
4467 4566: gap of unknown length  
4567 6225: contig of 1659 bp in length  
6226 6325: gap of unknown length  
6326 10630: contig of 4305 bp in length  
10631 10730: gap of unknown length  
10731 13945: contig of 3215 bp in length  
13946 14045: gap of unknown length  
14046 16402: contig of 2357 bp in length  
16403 16502: gap of unknown length  
16503 20172: contig of 3670 bp in length  
20173 20272: gap of unknown length  
20273 23395: contig of 3123 bp in length  
23396 23495: gap of unknown length  
23496 28819: contig of 5324 bp in length  
28820 28919: gap of unknown length  
28920 33023: contig of 4104 bp in length  
33024 33123: gap of unknown length  
33124 37026: contig of 3903 bp in length

\* 37027 37126: gap of unknown length  
\* 37127 41511: contig of 4385 bp in length  
\* 41512 41611: gap of unknown length  
\* 41612 47662: contig of 6051 bp in length  
\* 47663 53583: contig of unknown length  
\* 53584 53683: contig of 5821 bp in length  
\* 53684 63525: gap of unknown length  
\* 63526 63625: contig of 9842 bp in length  
\* 63626 73956: gap of unknown length  
\* 73957 73956: contig of 10231 bp in length  
\* 73957 81858: gap of unknown length  
\* 81859 81858: contig of 7702 bp in length  
\* 81859 94758: gap of unknown length  
\* 94759 94758: contig of 12720 bp in length  
\* 94759 108581: gap of unknown length  
\* 108582 108581: contig of 14003 bp in length  
\* 108582 127333: contig of unknown length  
\* 127334 127333: contig of 18652 bp in length  
\* 127334 145885: contig of unknown length  
\* 145886 145885: contig of 18452 bp in length  
\* 145886 175340: gap of unknown length  
\* 145986 175340: contig of 29355 bp in length.

FEATURES  
source

1. 175340  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15q21.3"  
/clone="CTD-3203H14"  
/clone.lib="Cal Tech human BAC library D"  
/note="This clone overlaps TP11-209E8 and RP11-6417"

BASE COUNT  
ORIGIN

53955 a 34442 c 32792 g 52011 t 2140 others  
Query Match 1.9% Score 21; DB 56; Length 175340;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 219 ttctccctccctccctgtcc 239  
|||||  
Db 105377 TTTCCCTCCCTCCCTGTCC 105357

RESULT 15  
AL354744/c

LOCUS  
DEFINITION

AL354744 183820 bp DNA HTG 29-JUL-2000  
Homo sapiens chromosome 9 clone RP11-635N21, \*\*\* SEQUENCING IN  
PROGRESS: \*\*\*, 14 unordered pieces.

ACCESSION  
VERSION

AL354744  
AL354744.4 GI:9588539

KEYWORDS  
SOURCE

HTG; HTGS\_PHASE1.  
human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 183820)

REFERENCE  
AUTHORS

Plumb, B.  
Direct Submission

TITLE  
JOURNAL

Submitted (28-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Jul 31, 2000 this sequence version replaced gi:9542773.

COMMENT

----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
----- Project Information  
Center project name: BA635N21  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2000, 22:07:51 ; Search time 3273.53 seconds  
(without alignments)  
2054.936 Million cell updates/sec

Title: US-09-101-423A-4  
Perfect score: 1088  
Sequence: 1 agttgggacacagctgtg.....acataataattttctgc 1088

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estom:\*  
68: em\_estov1:\*  
69: em\_estov2:\*  
70: em\_estpl1:\*  
71: em\_estpl2:\*  
72: em\_estpl3:\*  
73: em\_estpl4:\*  
74: em\_estpl5:\*  
75: em\_estpl6:\*  
76: em\_estro1:\*  
77: em\_estro2:\*  
78: em\_estro3:\*  
79: em\_estro4:\*  
80: em\_estro5:\*  
81: em\_estro6:\*  
82: em\_estro7:\*  
83: em\_estro8:\*  
84: em\_estro9:\*  
85: em\_estro10:\*  
86: em\_estro11:\*  
87: em\_estro12:\*  
88: gb\_gss1:\*  
89: gb\_gss2:\*  
90: gb\_gss3:\*  
91: gb\_gss4:\*  
92: gb\_gss5:\*  
93: gb\_gss6:\*  
94: gb\_gss7:\*  
95: gb\_gss8:\*  
96: gb\_gss9:\*  
97: gb\_gss10:\*  
98: gb\_gss11:\*  
99: gb\_gss12:\*  
100: gb\_gss13:\*  
101: gb\_gss14:\*  
102: gb\_gss15:\*  
103: gb\_gss16:\*  
104: gb\_gss17:\*  
105: gb\_gss18:\*  
106: gb\_gss19:\*  
107: gb\_gss20:\*  
108: gb\_gss21:\*  
109: gb\_gss22:\*  
110: gb\_gss23:\*  
111: gb\_gss24:\*  
112: gb\_gss25:\*  
113: gb\_gss26:\*  
114: gb\_gss27:\*  
115: gb\_gss28:\*  
116: gb\_gss29:\*

```

117: gb_gss18:*
118: gb_gss19:*
119: em_gss13:*
120: gb_gss20:*
121: gb_gss21:*
122: gb_gss22:*
123: gb_gss23:*
124: gb_gss24:*
125: em_gss14:*
126: em_gss15:*
127: em_gss16:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	58.4	5.4	581	117	AZ269259	RPCI-23-1
C 2	55.8	5.1	617	117	AZ228049	RPCI-23-9
	53.6	4.9	330	113	AQ993293	RPCI-23-3
3	51.2	4.7	573	115	AZ091764	RPCI-23-4
4	50.6	4.7	515	115	AQ086978	RPCI-23-4
5	49.4	4.5	504	14	AL024003	k8224b20
C 6	48.8	4.5	508	11	AI561895	v148b09.x
C 7	48.8	4.5	508	11	AI561895	v148b09.x
8	45.4	4.2	629	114	AZ086207	RPCI-23-4
9	43.4	4.0	614	114	AZ039772	RPCI-23-3
C 10	43.2	4.0	517	20	AW144209	EST294505
C 11	42.8	3.9	412	117	AZ262116	RPCI-23-1
C 12	42.6	3.9	822	123	CNS044CR	Tetraodon
C 13	41.6	3.8	543	114	AZ073260	RPCI-23-4
C 14	40.8	3.7	646	114	AZ044332	RPCI-23-3
15	40.8	3.7	724	113	AQ94907	RPCI-23-3
C 16	40.4	3.7	315	114	AZ040057	RPCI-23-2
C 17	39.6	3.6	362	20	AW143590	EST233795
18	39	3.6	384	2	AA1217999	msv57e03.r
C 19	38.4	3.5	379	9	AZ1207906	apl33f08.x
C 20	38.4	3.5	511	20	AW134993	UI-H-B11-
21	38	3.5	544	115	AZ106214	RPCI-23-4
C 22	38	3.5	637	117	AZ243908	RPCI-23-3
C 23	37.4	3.4	578	116	AZ224258	RPCI-23-7
C 24	36.6	3.4	284	30	BB391360	BB391360
25	36.6	3.4	531	90	AQ242626	HS_2193_A
26	36.4	3.3	697	123	CNS0485I	Tetraodon
27	35.8	3.3	287	33	BE110603	UI-R-BJ1-
28	35.4	3.3	290	29	BB325101	BB325101
C 29	35.4	3.3	315	30	BB357430	BB357430
C 30	35.4	3.3	481	34	BE220177	hV69801.x
31	35.2	3.2	395	6	AA724747	ah92h06.s
32	35.2	3.2	613	21	AW281193	fj51e09.y
33	35	3.2	285	30	BB343590	BB343590
34	35	3.2	290	24	AW819160	RC3-ST028
C 35	35	3.2	332	8	AI098994	uc01h03.r
36	35	3.2	395	11	AI615496	v134e05.y
37	35	3.2	430	5	AA561219	v134e05.r
C 38	35	3.2	437	5	AA982969	ua23h06.r
C 39	35	3.2	437	12	AI639814	ms85g12.y
C 40	35	3.2	466	2	AA170692	MS85g12.r
41	35	3.2	658	117	AZ253867	RPCI-23-5
42	35	3.2	977	121	CNS009H0	AL073588
43	34.8	3.2	296	26	BB031793	Drosophila
44	34.8	3.2	305	30	BB334621	BB334621
45	34.8	3.2	525	7	AA980588	ua43b02.r

## ALIGNMENTS

RESULT  
AZ269259

LOCUS	AZ269259	581 bp	DNA	GSS	26-JUL-2000				
DEFINITION	RPCI-23-160H10.TV RPCI-23 Mus musculus genomic clone RPCI-23-160H10 , DNA sequence.								
ACCESSION	AZ269259								
VERSION	AZ269259.1								
KEYWORDS	GSS.								
SOURCE	mouse house.								
ORGANISM	Mus musculus								
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
AUTHORS	1 (bases 1 to 581) Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret B., Levins, M., Megann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and de Fraser, C.N.M.								
TITLE	Mouse BAC End Sequences from Library RPCI-23								
JOURNAL	Unpublished (1999)								
COMMENT	Other-GSSs: RPCI-23-160H10.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.buffalo.edu). Clones may be purchased from BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu/orderingframe.htm">http://bacpac.med.buffalo.edu/orderingframe.htm</a> ) or from Reseta ch Genetics ( <a href="http://info@resgen.com">http://info@resgen.com</a> ). BAC end page: <a href="http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html">http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html</a> Plate: 150 row: H column: 10								

```

DHLUB electrocompetent cells (BRL Life Technologies).
195 a 107 c 140 g 139 t

BASE COUNT
ORIGIN

      5.4%; Score 58.4; DB 117; Length 581;
Best Local Similarity 53.9%; Pred. No. 1.8e-06;
Matches 192; Conservative 0; Mismatches 151; Indels 13; Gaps 3;

QY 740 tggcgccacaagcacagcgttttaaatctgtgggtgtgaccattatgactatcataa 799
|||||
Db 95 TGTATATGTAGACCAAGCTCTTTAAACAGTGAGTGTACCCCATTTGGGATG 154
|||||
QY 800 ctgagtgcagytatcaagaatactttagcagggtggtaaaaagatttttgaatgcgcaacg 859
|||||
Db 155 CTGAGTGTGGAGGTCAACAACAAGTTGACAACTG--TGAAGCTTCTTCAGCGGCACAACTG 212
|||||
QY 860 accaaactgaactcaaaaatcaagcatggcatggatccttgggtgctccttgggaagcactt 919
|||||
Db 213 ACCAAAAATATTTCAAGAACCAAACTTCGAATGAATCTGAGATGTGCTTGAAGTGCTCTC 272
|||||
QY 920 gccttttact-----gcattgtgcgactgtacggttagccttgggtttctgaatgcacaaca 972
|||||
Db 273 GTCTGTGCTTGCACAGTGTGATGCTGACCAACAACAACACTGGGCTCTGTGATGTCATCA 332
|||||

```







Db	66	GAATCGAGGTCTTCTGTGCCGAGCTTGCTGTATTACAGCAAAAAGTCTCCATGCCAGCC	7
QY	953	ttgggtt	958
Db	6	TTGGTT	1

RESULT	8
AZ086207	G29 bp DNA GSS 08-MAY-2000
LOCUS	RPCI-23-454M11.TJ RPCI-23 Mus musculus genomic clone RPCI-23-454M11
DEFINITION	, DNA sequence.
ACCESSION	AZ086207
VERSION	AZ086207.1 GI:7727941
KEYWORDS	GSS
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 629)
AUTHORS	Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret B., Levins,M., McGann,S., Tseayee,G., Geer,K., Krol,M., de Jong,P.F. and Fraser,C.M.
TITLE	Mus BAC End sequences from Library RPCI-23
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: RPCI-23-454M11.TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pietere@jeng.med.buffalo.edu). Clones may be purchased from BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu/orderingframe.htm">http://bacpac.med.buffalo.edu/orderingframe.htm</a> ) or from Resea ch Genetics ( <a href="mailto:info@resgen.com">info@resgen.com</a> ). BAC end page: <a href="http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html">http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html</a> Plate: 454 row: M column: 11 Seq primer: SP6 Class: BAC ends.

BASE COUNT	174 a	158 c	165 g	132 t
ORIGIN	DH10B electrocompetent cells (BKL Life Technologies)			

Query Match 4.2%; Score 45.4; DB 114; Length 629;  
Best Local Similarity 62.3%; Pred. No. 0.014;  
Matches 104; Conservative 0; Mismatches 61; Indels 2; Gaps 2;

Qy	872	ctcaaaaaatcaagcattggcattggaattctctctgggtgtctctctggaagcaact	918
Db	262	-TCTGAATCTAGATGGCGCAGTGAGTCCGAGGAGCGCGTGGCAGTGCT	307

RESULT	9	
AZ039772		
LOCUS	614 bp	DNA
DEFINITION	AZ039772.1 GI:7130569 Mus musculus genomic clone RPCI-23-365B14.TV RPCI-23 Mus musculus genomic clone RPCI-23-365B14	
ACCESSION	AZ039772	
VERSION	AZ039772.1	GI:7130569
KEYWORDS	mus mouse, mus musculus	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;	
REFERENCE	1 (bases 1 to 614)	
AUTHORS	Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Kraser, C.M.	
TITLE	Mouse BAC End Sequences from Library RPCI-23	
JOURNAL	Unpublished (1999)	
COMMENT	Other_GSSs: RPCI-23-365B14.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.buffalo.edu). Clones may be purchased from BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu/orderingframe.htm">http://bacpac.med.buffalo.edu/orderingframe.htm</a> ) or from Resea ch Genetics ( <a href="mailto:info@resgen.com">info@resgen.com</a> ). BAC end page: <a href="http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html">http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html</a> Plate: 355 row: B column: 14 Seq primer: T7 Class: BAC ends.	

BASE COUNT	170 i	125 c	135 g	184 t
ORIGIN	DH10B electrocompetent cells (BKL Life technologies)			

Query Match 4.0%; Score 43.4; DB 114; Length 514;  
Best Local Similarity 63.6%; Pred. No. 0.057;  
Matches 82; Conservative 0; Mismatches 46; Indels 1;  
Gaps 1;











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:11:30 ; Search time 141.08 seconds  
(without alignments)  
1166.364 Million cell updates/sec

Title: US-09-101-423A-4  
Perfect score: 1088  
Sequence: 1 agtggggacacagcttgct.....acataataattttctgc 1088

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2.6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2.6/ptodata/2/ina/6\_COMB.seq.\*  
6: /cgn2.6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
7: /cgn2.6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	39.8	3.7	7218	1	US-08-232-463-14
2	33.4	3.1	1923	1	US-08-077-939-14
3	33.4	3.1	1923	1	US-08-461-599-14
4	33.4	3.1	1923	1	US-08-461-621-14
5	33.4	3.1	1923	2	US-08-465-334-14
6	31.8	2.9	1974	1	US-08-077-939-16
7	31.8	2.9	1974	1	US-08-461-599-16
8	31.8	2.9	1974	1	US-08-461-621-16
9	31.8	2.9	1974	2	US-08-465-334-16
10	31	2.8	2049	1	US-08-268-797-1
11	31	2.8	2049	6	PCT-US95-08414-1
C 12	31	2.8	2094	5	US-08-714-918-87
13	31	2.8	2691	1	US-07-878-960-1
14	30.8	2.8	3523	5	US-08-749-527-1
15	30.6	2.8	3827	3	US-08-447-031A-1
16	30.6	2.8	4612	3	US-08-447-031A-8
17	30.6	2.8	9278	1	US-08-243-542-9
18	30.6	2.8	9278	1	US-08-477-407-9
19	30.6	2.8	9278	2	US-08-484-355-9
C 20	30.2	2.8	8224	4	US-09-010-398-14
21	30	2.8	3305	5	US-08-714-918-76
22	29.6	2.7	2820	2	US-08-286-305A-4
23	29.6	2.7	2820	3	US-08-441-104A-4
24	29.6	2.7	2820	4	US-08-440-816A-4
C 25	29.6	2.7	10367	1	US-08-110-300A-9
C 26	29.6	2.7	10367	4	US-08-886-642-9

C 27 29.6 2.7 10367 6 PCT-US93-08041-9 Sequence 9, Appli  
28 29.2 2.7 1985 5 US-08-619-812-1 Sequence 1, Appli  
C 29 29.2 2.7 2623 4 US-08-973-675-1 Sequence 1, Appli  
C 30 29 2.7 35100 2 US-08-306-691B-19 Sequence 19, Appli  
C 31 29 2.7 35100 6 PCT-US93-06251-19 Sequence 19, Appli  
C 32 28.6 2.6 1679 1 US-08-117-965-27 Sequence 27, Appli  
C 33 28.6 2.6 1766 6 PCT-US92-06532-2 Sequence 2, Appli  
34 28.6 2.6 40352 5 US-08-846-111B-15 Sequence 15, Appli  
35 28 2.6 2181 4 US-08-737-731A-1 Sequence 1, Appli  
36 28 2.6 2181 6 PCT-US95-05853-1 Sequence 1, Appli  
37 28 2.6 8323 1 US-08-110-300A-8 Sequence 8, Appli  
38 28 2.6 8323 4 US-08-886-642-8 Sequence 8, Appli  
39 28 2.6 8323 6 PCT-US93-08041-8 Sequence 8, Appli  
40 27.8 2.6 1716 4 US-08-832-028-15 Sequence 15, Appli  
C 41 27.8 2.6 2536 4 US-09-027-013-2 Sequence 2, Appli  
C 42 27.8 2.6 2536 5 US-09-244-233-2 Sequence 2, Appli  
43 27.8 2.6 3356 1 US-08-246-403A-1 Sequence 1, Appli  
44 27.8 2.6 3356 1 US-08-246-403A-4 Sequence 4, Appli  
C 45 27.8 2.6 3387 2 US-08-468-557-1 Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FAIKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 180C Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935.313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: ptz9pt-F1s  
US-08-232-463-14



Query Match : 3.1%; Score 33.4; DB 1; Length 1923;  
Best Local Similarity 55.7%; Pred. No. 0.56;  
Matches 64; Conservative 0; Mismatches 51; Indels 0

[illegible]

```

RESULT 5
US-08-465-334-14
; Sequence 14, Application US/08465334
; Patent No. 5728412
; GENERAL INFORMATION:
; APPLICANT: FUJII, Toshio
; APPLICANT: IWAMATSU, Akihiro
; APPLICANT: YOSHIMOTO, Hiroyuki
; APPLICANT: MINETOKI, Toshitaka
; APPLICANT: BOGAKI, Takayuki
; APPLICANT: NAGASAWA, Naoshi
; TITLE OF INVENTION: ALCOHOL
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,334  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/077,939  
FILING DATE: 18-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 62997/1993  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 184328/1992  
FILING DATE: 18-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 49441/101 KYPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1923 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear



Best Local Similarity 54.8%; Pred. No. 1.9; Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 754 acagcttttaaatctgtgggttgaccattatggaactatcataactgaagtcaggtat 813  
 Db 681 ACAAGAAATGAAGTGAAGTGGTGTCTCAATGAACAACCTGAGTACAGTACAGTAAT 740  
 QY 814 caagaataacttagcaggtggttaaaagatttttgatgcgaacgacacaaact 868  
 Db 741 GAAGCAATATAGAGAATTCAAAATAGTAGAGGGTTCCTATATCTGCAAAATTT 795

RESULT 8  
 US-08-461-621-16  
 ; Sequence 16, Application US/08461621  
 ; Patent No. 5686284  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FUJII, Toshio  
 ; APPLICANT: IWAMATSU, Akihiro  
 ; APPLICANT: YOSHIMOTO, Hiroyuki  
 ; APPLICANT: MINETOKI, Toshitaka  
 ; APPLICANT: BOGAKI, Takayuki  
 ; APPLICANT: NAGASAWA, Naoshi  
 ; TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington, D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/461,621  
 ; FILING DATE: 05-JUN-1995  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/077,939  
 ; FILING DATE: 18-JUN-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 62997/1993  
 ; FILING DATE: 26-FEB-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 184328/1992  
 ; FILING DATE: 18-JUN-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 49441/103 KYPA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)672-5300  
 ; TELEFAX: (202)672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 16:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1974 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 346..1923  
 ; US-08-461-621-16

Query Match 2.9%; Score 31.8; DB 1; Length 1974;  
 Best Local Similarity 54.8%; Pred. No. 1.9; Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 754 acagcttttaaatctgtgggttgaccattatggaactatcataactgaagtcaggtat 813  
 Db 681 ACAAGAAATGAAGTGAAGTGGTGTCTCAATGAACAACCTGAGTACAGTACAGTAAT 740  
 QY 814 caagaataacttagcaggtggttaaaagatttttgatgcgaacgacacaaact 868  
 Db 741 GAAGCAATATAGAGAATTCAAAATAGTAGAGGGTTCCTATATCTGCAAAATTT 795

RESULT 9  
 US-08-465-334-16  
 ; Sequence 16, Application US/08465334  
 ; Patent No. 5728412  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FUJII, Toshio  
 ; APPLICANT: IWAMATSU, Akihiro  
 ; APPLICANT: YOSHIMOTO, Hiroyuki  
 ; APPLICANT: MINETOKI, Toshitaka  
 ; APPLICANT: BOGAKI, Takayuki  
 ; APPLICANT: NAGASAWA, Naoshi  
 ; TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington, D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/465,334  
 ; FILING DATE: 05-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/077,939  
 ; FILING DATE: 18-JUN-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 62997/1993  
 ; FILING DATE: 26-FEB-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 184328/1992  
 ; FILING DATE: 18-JUN-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 49441/101 KYPA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)672-5300  
 ; TELEFAX: (202)672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 16:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1974 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 346..1923  
 ; US-08-465-334-16

Query Match 2.9%; Score 31.8; DB 2; Length 1974;  
 Best Local Similarity 54.8%; Pred. No. 1.9; Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
 QY 754 acagcttttaaatctgtgggttgaccattatggaactatcataactgaagtcaggtat 813.

Db 681 ARAAGATTGAACCTGAGTGGTGTCTCAATGACACCTGAGTACAGTGCAGTAAT 740  
Qy 814 caagaatactttgacagtggtgtaaaagatttttgaatgagcaacacaaact 868  
Db 741 GAAGCAAAATATTGAAGAATTCAAAATAGTAAGGGTTCTATACTGCAAAAT 795

## RESULT 10

US-08-268-797-1  
; Sequence 1, Application US/08268797  
; Patent No. 5599788  
; GENERAL INFORMATION:  
; APPLICANT: Purchio, Anthony F.  
; APPLICANT: LeBaron, Richard  
; TITLE OF INVENTION: Factor to Grow Tissue Ex Vivo  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson and Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/268,797  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned A.  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: TISSUE.001A  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2049 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; PCT-US-08-268-797-1

## Query Match 2.8%; Score 31; DB 1; Length 2049;

Best Local Similarity 51.9%; Pred. No. 3.5;  
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 442 gagactggccgggaacaaataaagtgagtgagtgatcatatgacagacacgtgagag 501  
Db 911 GAGACCTGCTGAACACACATCTTGAAGTCAGCTATGTGTGAGCCATCTGTTGCGG 970  
Qy 502 tcctgtccctagagatcataaaggacacgtgctggggactgttgagatggcactttg 561  
Db 971 GGCTGTCTGTAGACACCTGGAGGCGCACGACACTGGAGGTGGGCTGCAGCGGGACATGC 1030  
Qy 562 tgctgagagagggga 576  
Db 1031 TCACTATCAACGGGA 1045

## RESULT 11

PCT-US95-08414-1  
; Sequence 1, Application PC/TUS9508414  
; GENERAL INFORMATION:

; APPLICANT: Purchio, Anthony F.  
; APPLICANT: LeBaron, Richard  
; TITLE OF INVENTION: Factor to Grow Tissue Ex Vivo  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson and Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: PCT/US95/08414  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/268,797  
; FILING DATE: July 1, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned A.  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: TISSUE.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2049 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; PCT-US95-08414-1

## Query Match 2.8%; Score 31; DB 6; Length 2049;

Best Local Similarity 51.9%; Pred. No. 3.5;  
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 442 gagactggccgggaacaaataaagtgagtgagtgatcatatgacagacacgtgagag 501  
Db 911 GAGACCTGCTGAACACACATCTTGAAGTCAGCTATGTGTGAGCCATCTGTTGCGG 970  
Qy 502 tcctgtccctagagatcataaaggacacgtgctggggactgttgagatggcactttg 561  
Db 971 GGCTGTCTGTAGACACCTGGAGGCGCACGACACTGGAGGTGGGCTGCAGCGGGACATGC 1030  
Qy 562 tgctgagagagggga 576  
Db 1031 TCACTATCAACGGGA 1045

## RESULT 12

US-08-714-918-87/c  
; Sequence 87, Application us/08714918  
; Patent No. 6037123  
; GENERAL INFORMATION:  
; APPLICANT: Benzon, Bret  
; APPLICANT: Lee, Ving  
; APPLICANT: Malouin, Francois  
; APPLICANT: Martin, Patrick K.  
; APPLICANT: Schmid, Molly B.  
; APPLICANT: Sun, Dongxu  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL  
; TARGET GENES  
; NUMBER OF SEQUENCES: 111



;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Lyon & Lyon  
;; STREET: 633 West Fifth Street  
;; STREET: Suite 4700  
;; CITY: Los Angeles  
;; STATE: California  
;; COUNTRY: U.S.A.  
;; ZIP: 90071-2066  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
;; MEDIUM TYPE: storage  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: IBM P.C. DOS 5.0  
;; SOFTWARE: Word Perfect 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/714,918  
;; FILING DATE: September 13, 1996  
;; CLASSIFICATION: 424  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 60/009,102  
;; FILING DATE: December 22, 1995  
;; APPLICATION NUMBER: 60/003,798  
;; FILING DATE: September 15, 1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Warburg, Richard J.  
;; REGISTRATION NUMBER: 32,327  
;; REFERENCE/DOCKET NUMBER: 222/005  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (213) 489-1600  
;; TELEFAX: (213) 955-0440  
;; TELEX: 67-3510  
;; INFORMATION FOR SEQ ID NO: 87:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2094 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
US-08-714-918-87

Query Match 2.8%; Score 31; DB 5; Length 2094;  
Best Local Similarity 51.4%; Pred. No. 3.6;  
Matches 55; Conservative 6; Mismatches 46; Indels 0; Gaps 0;  
QY 699 cttatttcagtgaggtatttaccctgaggagaaagagctggtgccacaagcacgc 758  
Db 922 CTTCTTAACTGCGAGTAGTTATGATCTAATTGAATGCTGKGCAAAATWACTGA 863  
QY 759 tttaaatctgtggtgtgaccattatggactatcataactagct 805  
Db 862 TTYWMTCTTTGAATCTGCTAGTGTGGTATGATTGACGAGT 816

RESULT 13  
US-07-878-960-1  
; Sequence 1, Application US/07878960  
; Patent No. 544164  
; GENERAL INFORMATION:  
; APPLICANT: Purchio, Anthony F.  
; APPLICANT: Skonier, John  
; APPLICANT: Neubauer, Michael G.  
; TITLE OF INVENTION: TGF-BETA INDUCED GENE AND PROTEIN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bristol-Myers Squibb Company  
; STREET: 3005 First Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA: US/07/878,960  
;; FILING DATE: 05-MAY-1992  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/833,835  
;; FILING DATE: 05-FEB-1992  
;; NAME: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Sorrentino, Joseph M.  
;; REGISTRATION NUMBER: 32,598  
;; REFERENCE/DOCKET NUMBER: ON0092-  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 206/728-4800  
;; TELEFAX: 206/727-3601  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2691 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; HYPOTHETICAL: NO  
;; ORIGINAL SOURCE:  
;; ORGANISM: Hmo sapiens  
;; TISSUE TYPE: LUNG  
;; CELL TYPE: ADENOCARCINOMA  
;; CELL LINE: A549  
US-07-878-960-1

Query Match 2.8%; Score 31; DB 1; Length 2691;  
Best Local Similarity 51.9%; Pred. No. 4.1;  
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
QY 442 gagactggcctgggaacaaataaagtgaagtgaagtcataatgacagcacctgagag 501  
Db 958 GAGACTGCTGTAACACCCACATCTTGAAGTCAGCTATGTGCTGAACCCATCGTTGGG 1017  
QY 502 tctgtccctagatcataaggaacctggctgctgggactgtgttcagatggcactttg 561  
Db 1018 GGCTGTCTGTAGAGCCCTGGAGGCGCACACACTGGAGTGGGCTGCAGCGGACATGC 1077  
QY 562 tctcagagagagggga 576  
Db 1078 TCACATCATCAACGGGA 1092

RESULT 14  
US-08-749-527-1  
; Sequence 1, Application US/08749527  
; Patent No. 6054632  
; GENERAL INFORMATION:  
; APPLICANT: Reid, Marion E.  
; TITLE OF INVENTION: METHOD OF MAKING MONOCLONAL  
; TITLE OF INVENTION: ANTIBODIES USING POLYMORPHIC  
; TITLE OF INVENTION: TRANSGENIC ANIMALS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foffmann & Baron, LLP  
; STREET: 350 Jericho Turnpike  
; CITY: Jericho  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 11753  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WcrGperfect  
; CURRENT APPLICATION DATA:







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:18:59 ; Search time 254.92 Seconds  
(without alignments)  
1603.330 Million cell updates/sec

Title: US-09-101-423A-4

Perfect score: 1088

Sequence: 1 agtgggggacagcttgct.....acataataattttctgc 1088

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_36:\*

- 1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:\*
- 2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:\*
- 3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:\*
- 4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:\*
- 5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:\*
- 6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:\*
- 7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:\*
- 8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:\*
- 9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:\*
- 10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:\*
- 11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:\*
- 12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:\*
- 13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:\*
- 14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:\*
- 15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:\*
- 16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:\*
- 17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:\*
- 18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:\*
- 19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:\*
- 20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:\*
- 21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1088	100.0	1088	18	T72788 Metastasis inducin
2	33.4	3.1	1923	14	Q53305 Sake yeast alcohol
3	33.4	3.1	1923	15	077672 Sequence for contr
C 4	32.4	3.0	1140	20	Z23890 Human LOBO cDNA fr
C 5	32.4	3.0	1784	20	Z23898 Human LOBO homolog
C 6	32.2	3.0	335	16	T19127 Human gene signatu
C 7	31.8	2.9	1974	14	Q53306 Brewery yeast alco
C 8	31.8	2.9	1974	15	077673 Human gene signatu
C 9	31.6	2.9	1119	19	V15418 Sequence for contr
C 10	31.4	2.9	1664976	19	V21209 Human dendritic ce
C 11	31	2.8	332	16	T21637 Methanococcus jann
C 12	31	2.8	2049	17	T10137 H3 protein cDNA.

C 13	31	2.8	2094	21	A26936 Essential Staphylo
C 14	31	2.8	2691	14	O48043 TGF-beta induced g
C 15	31	2.8	2691	16	Q98815 Human beta-IG-H3 (
C 16	31	2.8	6531	20	X13012 Enterococcus faeca
C 17	30.8	2.8	3523	19	V11854 Human Duffy genomi
C 18	30.8	2.8	3523	19	V27017 Homo sapiens DNA f
C 19	30.8	2.8	5369	20	X78080 A. gossypii ADE4 D
C 20	30.6	2.8	424	20	V85976 EST clone BK112.
C 21	30.6	2.8	441	19	T93436 Collagen binding p
C 22	30.6	2.8	849	19	T93437 Collagen binding p
C 23	30.6	2.8	1215	19	V04470 Gallus gallus avia
C 24	30.6	2.8	1500	19	T93438 Collagen binding p
C 25	30.6	2.8	4612	13	Q24123 p16 and cCOLR6A as
C 26	30.6	2.8	9278	16	Q76124 Human MDC genomic
C 27	30.6	2.8	16125	20	Z21226 Human CGLCE genomi
C 28	30.6	2.8	144460	21	Z93815 Olfactory receptor
C 29	30.4	2.8	614	21	A16095 Human colon cancer
C 30	30.4	2.8	994	20	X40096 Gastric cancer ass
C 31	30.4	2.8	2029	21	Z52573 Human secreted pro
C 32	30.4	2.8	2784	18	T65102 Luciferase gene un
C 33	30.2	2.8	616	20	X61784 B. burgdorferi ant
C 34	30.2	2.8	690	20	X61783 B. burgdorferi ant
C 35	30.2	2.8	5805	20	X20268 Borrelia burgdorfe
C 36	30.2	2.8	8224	20	X80792 Semenogelin II gen
C 37	30.2	2.8	8224	21	A10424 Human semenogelin
C 38	30	2.8	969	20	Z07815 S. aureus nrdF pol
C 39	30	2.8	3305	21	A26925 Essential Staphylo
C 40	30	2.8	1230025	20	X91990 Nucleotide sequenc
C 41	30	2.8	1664976	19	V21209 Methanococcus jann
C 42	29.8	2.7	300	21	A00982 Human colon cancer
C 43	29.8	2.7	729	21	A02007 Human colon cancer
C 44	29.8	2.7	7646	18	V74432 Staphylococcus aur
C 45	29.8	2.7	32768	20	X12954 Enterococcus faeca

## ALIGNMENTS

RESULT 1

T72788

ID T72788 standard; DNA; 1088 BP.

XX

AC T72788;

XX

DT 22-SEP-1997 (first entry)

XX

DE Metastasis inducing DNA c9.

XX

KW Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;

KW osteopontin; ss.

XX

OS Homo sapiens.

XX

FN WO9725443-A1.

XX

PD 17-JUL-1997.

XX

PF 10-JAN-1997; 97WO-G500074.

XX

PR 10-JAN-1996; 96GB-0000470.

XX

PA (UYLI-) UNIV LIVERPOOL.

XX

PI Barraclough BR., Rudland PS;

XX

DR WPI; 1997-372878/34.

XX

PT New isolated metastasis-inducing DNA - used to develop products to identify and treat patients at risk from metastatic tumours

XX

PS Claim 11; Page 26; 38pp; English.

XX

CC Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12

CC and C20 (T72785-90) are entirely novel short stretches of human  
 CC regulatory DNA capable of inducing metastasis. They were  
 CC identified using a method for detecting Met-DNA that involves  
 CC transferring human DNA from a malignant, metastatic cancer cell  
 CC (in this case, breast cancer) into a cell line (pref. rat Rama 37)  
 CC that produces only benign, non-metastasizing tumours when injected  
 CC into a syngeneic animal, infecting the transformed cells into a  
 CC syngeneic animal, selecting those animals having metastasizing  
 CC tumours, and recovering the Met-DNA from them. The isolated  
 CC Met-DNAs can be used to develop products to identify and treat  
 CC patients at risk from metastatic tumours. Hybridisation of C9 DNA  
 CC occurred to HindIII-digested DNA from 4/9 human breast tumours  
 CC examined; there was no hybridisation to HindIII fragments from  
 CC healthy human breast or colon tissue.

XX Sequence 1088 BP; 265 A; 223 C; 302 G; 298 T; 0 other;

Query Match 100.0%; Score 1088; DB 18; Length 1088;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1088; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 agtggggacacagctgtgttgattgaagatgtttcttgggaagaggatgaagcctaag 60  
 Db 1 agtggggacacagctgtgttgattgaagatgtttcttgggaagaggatgaagcctaag 60

Qy 61 atttccaatgaaagagactcctaattggggaggcaatgttgccttaattgggacacactgcg 120  
 Db 61 atttccaatgaaagagactcctaattggggaggcaatgttgccttaattgggacacactgcg 120

Qy 121 ggttaattaaaagctctctccagtgcccttctctgttttggctctgggagcggaagagca 180  
 Db 121 ggttaattaaaagctctctccagtgcccttctctgttttggctctgggagcggaagagca 180

Qy 181 ttgagagggatgcaggcattcttaaggctgtgttctgttcttccctccctctgtcca 240  
 Db 181 ttgagagggatgcaggcattcttaaggctgtgttctgttcttccctccctctgtcca 240

Qy 241 aactcagtgaggatccctctgtctgtctccttagagtcctcctcctcctgagccttggtga 300  
 Db 241 aactcagtgaggatccctctgtctgtctccttagagtcctcctcctcctgagccttggtga 300

Qy 301 gtttaagctctctgagctgagctcctcagggaagcagtcagctcatttggaagggggag 360  
 Db 301 gtttaagctctctgagctgagctcctcagggaagcagtcagctcatttggaagggggag 360

Qy 361 aaccaggcaagggtgttggctgtgacctcagaaattcttgaggggcgaagggttcaaggctaa 420  
 Db 361 aaccaggcaagggtgttggctgtgacctcagaaattcttgaggggcgaagggttcaaggctaa 420

Qy 421 ctctcattatagacaagtttgagactggcctgggaacaaataataaagtgagtgaggt 480  
 Db 421 ctctcattatagacaagtttgagactggcctgggaacaaataataaagtgagtgaggt 480

Qy 481 catatgacagcactgaggagctcctccttagagatcataaggaacctgctcctgggga 540  
 Db 481 catatgacagcactgaggagctcctccttagagatcataaggaacctgctcctgggga 540

Qy 541 cttgttgagatggcactttgtgtcgagagagggaggaacctgcccagcatgggggacctg 600  
 Db 541 cttgttgagatggcactttgtgtcgagagagggaggaacctgcccagcatgggggacctg 600

Qy 601 gaagactcctctgattgaactgtgaactgttctcttatacctgaggtgtgctgt 660  
 Db 601 gaagactcctctgattgaactgtgaactgttctcttatacctgaggtgtgctgt 660

Qy 661 tatctgttacacatctctgtgggtaagtgagttcattggcctttatttcagtgaggtattta 720  
 Db 661 tatctgttacacatctctgtgggtaagtgagttcattggcctttatttcagtgaggtattta 720

Qy 721 cctgagagagaagagactgtgtcccaagacacagcttttaaatctgtgggtgtgac 780  
 Db 721 cctgagagagaagagactgtgtcccaagacacagcttttaaatctgtgggtgtgac 780

Qy 781 ccattatggactatcataactgagtgagggtatcaagaatactttagcaggtgtgtaaaaa 840  
 Db 781 ccattatggactatcataactgagtgagggtatcaagaatactttagcaggtgtgtaaaaa 840

Qy 841 gatttttgaa:gcgcaacgacccaactgaactcaaaaaatacaagcatggcagtgatcctg 900  
 Db 841 gatttttgaa:gcgcaacgacccaactgaactcaaaaaatacaagcatggcagtgatcctg 900

Qy 901 ggtgctcctgaaagcaacttgcctttactgactgtgcgacttgacggtagccttggttct 960  
 Db 901 ggtgctcctgaaagcaacttgcctttactgactgtgcgacttgacggtagccttggttct 960

Qy 961 gaatgcacaacacgtggcgtttgggtgcacagggccaccacgcctgctcctgaaacacctc 1020  
 Db 961 gaatgcacaacacgtggcgtttgggtgcacagggccaccacgcctgctcctgaaacacctc 1020

Qy 1021 agctcaggtt:gtggctatgtcctatgacttgacttgacttacttttattgacataataat 1080  
 Db 1021 agctcaggtt:gtggctatgtcctatgacttgacttgacttacttttattgacataataat 1080

Qy 1081 ttctctgc 1088  
 Db 1081 ttctctgc 1088

## RESULT 2

Q53305  
 ID Q53305 standard; DNA; 1923 BP.

XX Q53305;

AC AC Q53305;

DT 17-JUN-1994 (first entry)

XX Sake yeast alcohol acetyl transferase gene.

XX AATase; Saccharomyces cerevisiae; sake yeast; alcohol production;  
 KW fermentation; alcoholic beverage; enriched ester flavour; ss.

XX Saccharomyces cerevisiae (Kyokai No.7).

XX Key Location/Qualifiers

FT CDS 234..1811

FT /\*tag= a

FT /product= AATase

XX EP574941-A.

XX 22-DEC-1993.

XX 18-JUN-1993; 33EP-0109775.

XX 18-JUN-1992; 32JP-0184328.

XX 26-FEB-1993; 33JP-0062997.

XX (KIRI ) KIRIN BEER KK.

XX (KIRI ) KIRIN BREWERY KK.

XX (OZEK-) OZEKI KK.

XX Bogaki T, Fujil T, Iwamatsu A, Minetoki T, Nagasawa N;

XX Yoshimoto H;

XX WPI; 1993-407311/51.

XX P-PSDB; R44825;

XX New alcohol acetyl transferase(s) with specified sequences

XX derived from yeast - useful for transforming yeast to produce

XX alcoholic beverage with enriched ester flavour

XX Claim 10; Fig 1: 38pp; English.

XX An AATase was isolated from sake yeast (Kyokai #7), purified and

XX partially sequenced. Two degenerate probe pools were designed based

CC



[illegible]



Db 64 GTGCTGTTAAGATAGAACATATTACAC 36

RESULT 7

Q53306  
ID Q53306 standard; DNA; 1974 BP.  
XX AC Q53306;  
XX DT 17-JUN-1994 (first entry)  
XX DE  
XX DE Brewery yeast alcohol acetyl transferase 1 gene.

XX ATase; Saccharomyces cerevisiae; Brewer's yeast; Lager yeast;  
XX alcohol production; fermentation; alcoholic beverage;  
XX enriched ester flavour; ss.

XX OS Saccharomyces cerevisiae.

XX FH Key Location/Qualifiers  
XX CDS 346..1923  
XX FT /\*tag= a  
XX FT /product= AATase\_1

XX PN EP574941-A.  
XX PD 22-DEC-1993.  
XX PF 18-JUN-1993; 93EP-0109775.  
XX PR 18-JUN-1992; 92JP-0184328.  
XX PR 26-FEB-1993; 93JP-0062997.

XX PA (KIRI ) KIRIN BEER KK.  
XX PA (KIRI ) KIRIN BREWERY KK.  
XX PA (OZEK-) OZEKI KK.

XX PI Bogaki T, Fujii T, Iwamatsu A, Minetoki T, Nagasawa N;  
XX PI Yoshimoto H;  
XX DR WPI; 1993-407311/51.  
XX DR P-PSDB; R44826.

XX PT New alcohol acetyl transferase(s) with specified sequences  
XX PT derived from yeast - useful for transforming yeast to produce  
XX PT alcoholic beverage with enriched ester flavour

XX PS Claim 10; Fig 2; 38pp; English.  
XX CC The AATase gene from sake yeast was used as a probe to isolate  
XX CC AATase coding sequences from Brewery lager yeast. Hybridising  
XX CC clones fell into two groups, representing Brewer's yeast AATase 1  
XX CC and AATase 2 enzymes. The DNA sequences of the two genes correspond  
XX CC to Q53306 and Q53307, respectively. Yeast transformed with these  
XX CC will be useful for producing alcoholic beverages with an enriched  
XX CC ester flavour. See also Q53305 for the sake yeast AATase gene.  
XX SQ Sequence 1974 BP; 614 A; 364 C; 386 G; 610 T; 0 other;

Query Match 2.9%; Score 31.8; DB 14; Length 1974;  
Best Local Similarity 54.8%; Pred. No. 5.1;  
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 754 acagcttttaactgtgggttgaccattatggactatcataactgagtcaggtat 813  
Db 681 acaagaattgaactgagtggtgttctcaatgaacacctgagtcagtcagta 740  
Qy 814 caagaatacttagcaggtgtgtaaaagattttgaaatgcgaacgacaaaact 868  
Db 741 gaagcaaatatagaagaattcaaaaatagtaagggttctctactgcaaaaatt 795

RESULT 8

Q77673  
ID Q77673 standard; DNA; 1974 BP.  
XX AC Q77673;  
XX DT 21-JUL-1995 (first entry)  
XX DE  
XX DE Sequence for controlling alcohol acetyl transferase expression.  
XX Yest; alcohol acetyl transferase; AATase; inhibition; regulation;  
XX KW ester-flavour reduced spirits; ss.

XX OS Saccharomyces cerevisiae.

XX FH Key Location/Qualifiers  
XX CDS 346..1923  
XX FT /\*tag= a

XX PN JP06253826-A.  
XX PD 13-SEP-1994.  
XX PF 26-FEB-1993; 93JP-0062998.  
XX PR 26-FEB-1993; 93JP-0062998.  
XX PA (KIRI ) KIRIN BREWERY KK.  
XX DR WPI; 1994-328981/41.  
XX DR P-PSDB; R63190.

XX PT Transformed yeast where alcohol acetyl-transferase activity is  
XX PT reduced or inhibited - is useful for prodn. of ester-flavour reduced  
XX PT spirits obtd. by sugar fermentation using the yeast

XX PS Claim 1; Fig 3 and 4; 27pp; Japanese.

XX CC The expression of alcohol acetyl-transferase enzyme (AATase) is  
XX CC controlled using the coding regions of Q77672, Q77673 or Q77674.  
XX CC Yeast in which the AATase activity is reduced or inhibited are  
XX CC useful for production of ester flavour-reduced spirits.

XX SQ Sequence 1974 BP; 614 A; 364 C; 386 G; 610 T; 0 other;

Query Match 2.9%; Score 31.8; DB 15; Length 1974;  
Best Local Similarity 54.8%; Pred. No. 5.1;  
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 754 acagcttttaactgtgggttgaccattatggactatcataactgagtcaggtat 813  
Db 681 acaagaattgaactgagtggtgttctcaatgaacacctgagtcagtcagta 740  
Qy 814 caagaatacttagcaggtgtgtaaaagattttgaaatgcgaacgacaaaact 868  
Db 741 gaagcaaatatagaagaattcaaaaatagtaagggttctctactgcaaaaatt 795

RESULT 9

V15418/C  
ID V15418 standard; CDNA; 1119 BP.  
XX AC V15418;  
XX DT 11-JUN-1998 (first entry)

XX DE Human dendritic cell chemokine receptor encoding cDNA.  
XX KW Human; thymus expressed chemokine; TECK; MIP-3alpha; MIP-3beta;  
XX KW receptor; dendritic cell; macrophage; inflammation; asthma; ss.

OS Homo sapiens.  
 XX Key Location/Qualifiers  
 XX CDS 1..1098  
 FT /\*tag= a  
 FT /product= "DC CR"  
 FT /note= "dendritic cell chemokine receptor"  
 FT 577..579  
 FT /\*tag= b  
 FT /note= "encodes His or Gln, but is stated as Gln in  
 FT the protein (shown in W48086)."  
 XX W09801557-A2.  
 XX PN  
 XX PD 15-JAN-1998.  
 XX PF 02-JUL-1997; 97WO-US10819.  
 XX PR 04-JUN-1997; 97US-0048593.  
 XX PR 05-JUL-1996; 96US-0675814.  
 XX PR 11-OCT-1996; 96US-0028329.  
 XX SCHE ) SCHERING CORP.  
 XX Gish KC, Schall TJ, Vicari A, Wang W, Zlotnik A;  
 WPI: 1998-101054/09.  
 P-PSDB: W48086.  
 Novel chemokines, e.g. thymus expressed chemokine - used for  
 treating inflammatory conditions including asthma.  
 Claim 5; Page 92-94; 202pp; English.  
 The present sequence encodes a human dendritic cell chemokine receptor.  
 Antibodies which bind to the protein can be used in detecting or  
 diagnosing various immunological conditions related to expression  
 of the protein. The nucleic acid can be used for screening and  
 isolating DNA clones for the chemokines, especially from other  
 species. The chemokine can be used in the treatment of conditions  
 associated with abnormal physiology or development, including  
 inflammatory conditions such as asthma.  
 Sequence 1119 BP; 261 A; 268 C; 257 G; 332 T; 1 other;  
 Query Match 2.9%; Score 31.6; DB 19; Length 1119;  
 Best Local Similarity 54.2%; Pred. No. 4.3;  
 Matches 64; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
 Qy 327 tcagggaacgcagtcgctgctggaaggagggaacccaggcaagggtgtgctgtgac 386  
 Db 1119 TTAGGAGACTCAGCTTCTATCATAGTAGAGGAGCGCATTCGCTTATCTGCGGT 1060  
 Qy 387 ctccagaattctgaggggcaagggttcagggttaactctattatagagcaagtttag 444  
 Db 1059 CTCAGTGTCTGCGGAGAAATGTTTTCTGAGTACCTCCCGGCACAGGAGAGCTGAG 1002  
 RESULT 10  
 V21209  
 ID V21209 standard; DNA; 1664976 BP.  
 XX AC V21209;  
 XX DT 10-NOV-1998 (first entry)  
 XX DE Methanococcus jannaschii circular chromosome.  
 XX KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;  
 XX genome; autotrophic; extrachromosomal element; identification; ds.  
 XX OS Methanococcus jannaschii.

XX W09807830-A2.  
 XX PN  
 XX PD 26-FEB-1998.  
 XX PF 22-AUG-1997; 97WO-US14900.  
 XX PR 22-AUG-1996; 96US-0024428.  
 XX (GENO-) INST GENOMIC RES.  
 XX (UNII ) UNIV ILLINOIS FOUND.  
 XX (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 XX Bult CJ, Smith HO, Venter JC, White OR, Woese CR;  
 WPI: 1998-169143/15.  
 Complete genome sequence of methano-genic archaeon, Methanococcus  
 jannaschii - useful in identification of M. jannaschii genome  
 fragment  
 Claim 13; Page 152-585; 614pp; English.  
 The present sequence represents the complete 1.66-megabase pair genome  
 sequence of the Methanococcus jannaschii circular chromosome. The  
 present invention describes M. jannaschii open reading frames from the  
 genome sequence. The invention also describes a computer based system  
 for identifying fragments of the M. jannaschii genome that are  
 homologous to target nucleotide sequences, comprising: (a) data storage  
 means comprising the nucleotide sequence of the 1664976, 58407 or 16550  
 bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at  
 least 99.9% identical to it; (b) search means for comparing a target  
 sequence to the nucleotide sequence of the data storage means to  
 identify a homologous sequence, and (c) retrieval means for obtaining  
 the homologous sequence. The method, which is based on whole genome  
 random sequencing of an autotrophic archaeon M. jannaschii, the genome  
 of which consists of 3 physically distinct elements, a large circular  
 chromosome (the 1664976 bp sequence given in V21209), a large circular  
 extra-chromosomal element (the 58407 bp sequence given in V21210), and a  
 small circular extra-chromosomal element (the 16550 bp sequence given in  
 V21211), can be used in the identification of M. jannaschii genome  
 fragment.  
 Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;  
 Query Match 2.9%; Score 31.4; DB 19; Length 1664976;  
 Best Local Similarity 57.7%; Pred. No. 1.6e+02;  
 Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
 Qy 683 tgaatgagttcatggcgtttatttcagtgaggtatttacctgaggaagaagactgg 742  
 Db 296160 tgcagcgtttaaggccattagattgattgattgtctctggaccctaaaggggatatg 296219  
 Qy 743 tgcacacaaacacagcttttaaatctgtgggttgta 779  
 Db 296220 tgcacacaaacaaagcttatagagctttgggattaga 296256  
 RESULT 11  
 T21637/c  
 ID T21637 standard; cDNA to mRNA; 332 BP.  
 XX AC T21637;  
 XX DT 06-AUG-1996 (first entry)  
 XX DE Human gene signature HUMGS03073.  
 XX KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 XX human; cloning; mapping; non-biased library; diagnosis; detection;  
 XX cell typing; abnormal cell function; ss.  
 XX



```

XX Benton B, Lee VJ, Martin PK, Schmid MB, Sun D, Malouin F;
XX WPI; 2000-282222/24.
XX
XX Novel methods for screening for antibacterial agents, useful for
XX treatment or prevention of Staphylococcus infection, by testing
XX compounds against, essential bacterial genes -
XX
XX Disclosure; Column 201-204; 256pp; English.
XX
XX The present sequence represents an essential Staphylococcus aureus gene.
XX The invention relates to a method for screening for an antibacterial
XX agent. The method comprises determining if a test compound is active
XX against the essential bacterial genes (A26850-A26956). The invention also
XX includes a method for evaluating an agent for activity on the essential
XX genes. Also included in the invention is the production of an
XX antibacterial agent. The antibacterial agents of the invention are used
XX to treat or prevent bacterial infections, particularly where caused by
XX Staphylococcus aureus. The antibacterial agents are unlikely to be
XX affected by known resistance mechanisms.
XX
XX Sequence 2094 BP; 725 A; 327 C; 420 G; 597 T; 25 other;

Query Match      2.8%; Score 31; DB 21; Length 2094;
Best Local Similarity 51.4%; Pred. No. 9.6;
Matches 55; Conservative 6; Mismatches 46; Indels 0; Gaps 0;

QY 599 cttatttcagtgaggatttacctgaggagaaagaggactgtgtcccaagacagc 758
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 922 CCTCTTAACTGAGTAGTAAATGATCATCTAATTAATGATGCTGKGCAACAATWACTGA 863

QY 759 ttttaactgtggtgtgaccattatgactatcataactgagt 805
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 862 TTYAWTCTTTGAATCTGGTAGTGTGTGTAATGATGATGACCAAGT 816

RESULT 14
Q48043
ID Q48043 standard; DNA; 2691 BP.
XX
AC Q48043;
XX
XX 07-FEB-1994 (first entry)
XX
XX TGF-beta induced gene betaIG-H3.
XX
XX Transforming growth factor-beta; cell growth; differentiation;
XX Arg-Gly-Asp; RGD; cell surface recognition; communication;
XX fascilin-I; transmission; intracellular signal; integrin;
XX gene regulation; ss.
XX
XX Homo sapiens.
XX
XX Key      Location/Qualifiers
XX CDS      48..2099
XX          /*tag= a
XX          /note= "beta IG-H3"
XX
XX sig_peptide 48..89
XX          /*tag= b
XX
XX misc_feature 89..90
XX          /*tag= c
XX
XX misc_feature 116..117
XX          /note= "cleavage site"
XX
XX repeat_unit 462..869
XX          /tag= d
XX          /note= "cleavage site"
XX
XX          /*tag= e
XX          /number= 1
XX
XX repeat_unit 870..1277
XX          /*tag= f
XX          /number= 2
XX

```

```

FT repeat_unit 1278..1658
FT /*tag= g
FT /number= 3
FT repeat_unit 1659..2096
FT /*tag= h
FT /number= 4
FT misc_feature 1971..1979
FT /*tag= i
FT /note= "RGD sequence"
FT polyA_signal 2672..2677
FT /*tag= j
XX
XX EP555989-A.
XX
XX 18-AUG-1993.
XX
XX 04-FEB-1993; 33EP-0300809.
XX
XX 05-FEB-1992; 32US-0833835.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Neubauer MG, Purchio AF, Skonier JE;
XX
XX WPI; 1993-260238/33.
XX P-PSDB; R40386.
XX
XX TGF-beta induced gene and protein - is designated BIG-H3 contg. 4
XX homologous repeats and possibly involved in cell surface
XX recognition and communication
XX
XX Claim 10; Fig 5: 23pp; English.
XX
XX The betaIG-H3 protein contains 4 homologous repeat regions (at least
XX 16% homology). The protein is possibly a novel cell surface
XX recognition protein (similar to fascilin-I) and may thus have a role
XX in cell-cell communication/transmission of intracellular signals
XX involved in negative growth control. In addition, the protein
XX contains Arg-Gly-Asp in the C-terminal which may allow interaction
XX with integrins. The protein may be useful in the study of TGF-beta
XX partic. its role in gene regulation.
XX
XX Sequence 2691 BP; 679 A; 729 C; 695 G; 588 T; 0 other;

Query Match      2.8%; Score 31; DB 14; Length 2691;
Best Local Similarity 51.9%; Pred. No. 11;
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 442 gagactggccgggaacacaaataaaagtgaagtgcattatgacagcacctgaggag 501
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 958 gagacctgctgaacacacacacattcttgagtcagctatgtgtgtaagccatggttcgg 1017

QY 502 tctgtccctagagatcataaggacctggctgctgggggactgttgcagatggcaactttg 561
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1018 ggctgtctgtagagacctggaggggcagcacactggagggtgggctgcaggggacatgc 1077

QY 562 tgtcgagagagggga 576
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1078 tcactatacaaggga 1092

RESULT 15
Q98815
ID Q98815 standard; DNA; 2691 BP.
XX
XX Q98815;
XX
XX 06-MAR-1996 (first entry)
XX
XX Human beta-IG-H3 (transforming growth factor-beta induced gene-h3).
XX
XX beta-IG-H3; antitumor; ss.

```

```

XX OS Homo sapiens.
XX PH Key
XX FT CDS
XX FT Location/Qualifiers
XX FT 48..2099
XX FT /tag= a
XX FT sig_peptide
XX FT 48..89
XX FT /tag= b
XX FT misc_feature
XX FT 89..90
XX FT /tag= c
XX FT /note= "cleavage site"
XX FT 116..117
XX FT /tag= d
XX FT /note= "cleavage site"
XX FT 2669..2677
XX FT /tag= e
XX FT repeat_region
XX FT 462..869
XX FT /tag= f
XX FT /note= "REPEAT 1"
XX FT 870..1277
XX FT /tag= g
XX FT /note= "REPEAT 2"
XX FT 1278..1658
XX FT /tag= h
XX FT repeat_region
XX FT 1659..2693
XX FT /tag= i
XX FT /note= "REPEAT 4"
XX PN
XX US5444164-A.
XX PD
XX 22-AUG-1995.
XX XX
XX 05-FEB-1992; 92US-0833835.
XX XX
XX 04-MAY-1992; 92US-0878960.
XX DR
XX 05-FEB-1992; 92US-0833835.
XX XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX XX
XX Neubauer MG, Purchio AF, Skonier JE;
XX PI
XX WPI; 1995-302161/39.
XX DR
XX P-PSDB; R80573.
XX XX
XX New isolated nucleic acid and protein, beta IG-H3 - which are
XX PT induced by mammalian cells treated with TGF-beta, used for
XX PT inhibiting growth of tumour cells
XX PT
XX PS Claim 1; Page 11-14; 23pp; English.
XX XX
XX DNA encoding betaHG-H3 was isolated from a cDNA library prepared
XX CC from A549 human lung adenocarcinoma cells treated with transforming
XX CC growth factor-beta. This sequence encodes a beta-IG-H3 protein, the
XX CC induction, expression and/or secretion of which inhibits the growth,
XX CC colony formation and appearance of tumor cells. The beta-IG-H3 can
XX CC be used to inhibit the growth of tumor cells.
XX XX
XX Sequence 2691 BP; 679 A; 729 C; 695 G; 588 T; 0 other;

```

```

Query Match 2.8%; Score 31; DB 16; Length 2691;
Best Local Similarity 51.9%; Pred. No. 11;
Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 442 gagactggcctgggaacaaataataaagtgaagtgaagtcataacacagcaccctgagag 501
Db ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
958 gagacctgtgaacacacacacacacacacacacacacacacacacacacacacacacac 1017
QY 502 tccgtccctagatcataagaccctgctgctgggggactgtgacagatggcactttg 561
Db ||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||
1018 ggcgtgtctagagaccctggaggacacacacacacacacacacacacacacacacacac 1077

```

```

QY 562 tgtcagagagagggga 576
Db 1078 tcaactatcaecggga 1092

```

Search completed: Nvember 4, 2000, 00:29:46  
Job time: 21456 sec.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: Nov 04, 2000, 00:54:09 ; Search time 7109.34 Seconds  
(without alignments)  
568.406 Million cell updates/sec

Title: US-09-101-423A-4  
Perfect score: 1088  
Sequence: 1 agtggggacacagcttgct.....acataataattttctgc 1088

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues  
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

##### GenEmbl:\*

1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pr1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: em\_fun.\*  
13: em\_hum1.\*  
14: em\_hum2.\*  
15: em\_in.\*  
16: em\_om.\*  
17: em\_or.\*  
18: em\_ov.\*  
19: em\_pat.\*  
20: em\_ph.\*  
21: em\_pl.\*  
22: em\_ro.\*  
23: em\_sts.\*  
24: em\_sy.\*  
25: em\_un.\*  
26: em\_vi.\*  
27: gb\_htg1.\*  
28: gb\_htg2.\*  
29: gb\_in1.\*  
30: gb\_in2.\*  
31: em\_ba1.\*  
32: em\_ba2.\*  
33: em\_hum3.\*  
34: em\_hum4.\*  
35: gb\_pr4.\*  
36: gb\_htg3.\*  
37: gb\_htg4.\*  
38: gb\_htg5.\*  
39: gb\_htg6.\*  
40: gb\_htg7.\*  
41: em\_htg1.\*  
42: em\_htg2.\*  
43: em\_htg3.\*

44: em\_hum5.\*  
45: gb\_pl3.\*  
46: gb\_pr5.\*  
47: gb\_htg8.\*  
48: gb\_htg9.\*  
49: gb\_htg10.\*  
50: gb\_htg11.\*  
51: gb\_htg12.\*  
52: gb\_htg13.\*  
53: gb\_htg14.\*  
54: gb\_in3.\*  
55: gb\_htg15.\*  
56: gb\_htg16.\*  
57: gb\_htg17.\*  
58: em\_htg4.\*  
59: em\_htg5.\*  
60: em\_htg6.\*  
61: em\_htg7.\*  
62: em\_hum6.\*  
63: gb\_htg18.\*  
64: gb\_htg19.\*  
65: gb\_ba3.\*  
66: em\_htg8.\*  
67: em\_htg9.\*  
68: em\_htg10.\*  
69: em\_htg11.\*  
70: em\_htg12.\*  
71: em\_htg13.\*  
72: em\_htg14.\*  
73: em\_htg15.\*  
74: em\_htg16.\*  
75: em\_htg17.\*  
76: em\_htg18.\*  
77: em\_htg19.\*  
78: em\_htg20.\*  
79: em\_htg21.\*  
80: em\_htg22.\*  
81: em\_htg23.\*  
82: gb\_pr6.\*  
83: gb\_pr7.\*  
84: gb\_htg20.\*  
85: gb\_htg21.\*  
86: gb\_htg22.\*  
87: gb\_htg23.\*  
88: gb\_ro.\*  
89: gb\_sts1.\*  
90: gb\_sts2.\*  
91: gb\_sy.\*  
92: gb\_un.\*  
93: gb\_vil.\*  
94: gb\_vil2.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1088	100.0	1088	5	A64061
C 2	148.6	13.7	171370	9	AC004021 Human PAC
C 3	91.8	8.4	405	3	AF185100 Lasiorhin
C 4	60.4	5.6	200094	47	AC020862
C 5	52.6	4.8	236056	56	AC069274 Mus muscu
C 6	51.4	4.7	143556	56	AC068907 Mus muscu
7	48.8	4.5	226798	48	AC022235 Mus muscu
8	47.6	4.4	24035	88	MMU251155
C 9	46.6	4.3	189308	57	AC073437
C 10	46.6	4.3	193816	50	AC025874 Mus muscu
11	46.6	4.3	207160	39	AC016464 Mus muscu
12	45	4.1	193260	50	AC025598 Homo sapi

```

13 44.6 4.1 193816 50 AC025874
14 44.4 4.1 187108 51 AC026681
15 44.2 4.1 110000 48 AC023048_2
16 44.2 4.1 189682 47 AC021060
17 43 4.0 133094 36 AC021160
18 42 3.9 186812 50 AC025382
19 39.8 3.7 7218 5 I66494
20 39.4 3.6 190004 47 AC020787
21 38.8 3.6 157225 57 AC074359
22 38.8 3.6 212896 50 AC024608
23 38.6 3.5 173359 50 AC025583
24 38.4 3.5 1200 82 HSCACNG4
25 38.4 3.5 169045 9 AC005544
26 38.2 3.5 466331 48 AC022518
27 38.2 3.5 176657 47 AC021751
28 38 3.5 110312 88 AF131866
29 38 3.5 152210 48 AC022910
30 38 3.5 206100 56 AC068651
31 38 3.5 240622 57 AC073793
32 37.8 3.5 197656 57 AC073600
33 37.6 3.5 1255 82 HSCRL3
34 37.6 3.5 1518 82 HSU68032
35 37.6 3.5 2978 82 HSU68030
36 37.6 3.5 3693 82 HSU45984
37 37.6 3.5 101230 82 HSPA517H2
38 37.6 3.5 178133 64 AL353747
39 37.4 3.4 19185 9 AB042425
40 37.4 3.4 36824 8 AP000372
41 37.2 3.4 37167 51 AC026806
42 37.2 3.4 41613 9 AC005625
43 37.2 3.4 180932 56 AC068627
44 37 3.4 107104 46 HS1179L24
45 37 3.4 186582 85 AL357561

```

## ALIGNMENTS

```

RESULT 1
LOCUS A64061 1088 bp DNA PAT 29-MAR-1999
DEFINITION Sequence 4 from Patent WO9725443.
ACCESSION A64061
VERSION A64061.1 GI:3717535
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1088)
AUTHORS Rudland,P.S. and Barracough,B.R.
TITLE METASTASIS INDUCING DNA'S
JOURNAL Patent: WO 9725443-A 17-JUL-1997;
UNIV LIVERPOOL (GB)
FEATURES
    source      Location/Qualifiers
    1..1088
    /organism="unidentified"
    /db_xref="taxon:32644"
BASE COUNT 265 a 223 c 302 g 298 t
ORIGIN
Query Match 100.0%; Score 1088; DB 5; Length 1088;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1088; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 agtggggacacagctgctgattgaatgtttctgggaaagagtgtaagcctaagc 60
Db 1 AGTGGGGACACAGCTGCTGATTGAATGTTCCTGGGAAAGAGTGAAGCCCTAATG 60
QY 61 atttccaatggaaagactgctaatgtgggagggcaatgtgtcttaattgggacacctgcg 120
Db 61 ATTTCCAATGGAAGACTGCTAATGTGGGAGGCAATGTGCTTAATTGGGACACCTCGC 120

```

```

QY 121 ggttaataaangctctctccagctggcccttccctgtgttttggctctgggagcgaaagca 180
Db 121 GGTAAATTAAGCTCTCTCCAGTGGCCCTTTCTCTGTTTGGCTCTGGGAGCCGAGGCA 180
QY 181 ttgagagggga-gcagaggaattctaaggcggtgtcttcttggttctccctccctctgtcca 240
Db 181 TTGAGAGGGGATGCAGGCAATCTTAAGGGCTGGTCTTGGTTCTCCCTCCCTCTGTCCA 240
QY 241 aactcagtgagtgatccctctctctgtctgtctcttagagtagccctcagcctgagctgtgta 300
Db 241 AACTCAGTGAGTGATCCCTCTGTCTGTCTTGTAGAGTCCCTGAGGCTTGTGTGA 300
QY 301 gftaagggtctctggatctgagctgcctcagggaaacacatgagctcatttgaaaggggag 360
Db 301 GTTAAGGTCTCTGGATCTGAGCTGCCCTCAGGGAACGCATGAGCTCATTTGGAAGGGGAG 360
QY 361 aaccagggcaagggtgtgtgctgtgacctcagaattctgaggggcaagggttcaagggttaa 420
Db 361 AACCAAGGCAAGGTTGTGGCTGTGACCTCAGATTTCTGAGGGGCAAGGTTCAAGGGTAA 420
QY 421 ctctcattatagacaagtgttgacctggcctggggaacaaataataaagtgaagtgggt 480
Db 421 CTCCTATTATAGACAAGTGTGAGACTGGCCCTGGGAACAAAATATAAAGTGAGTGAGGT 480
QY 481 catatgacag-acacctgagggagctcctgtccctagatgacataagacacctggctgctgggga 540
Db 481 CATATGACAGACCTGAGGAGTCTCTGCTCCCTAGAGATCATTAAGGACCTGGCTGCTGGGA 540
QY 541 cttgttcagatggcaactttgtctgagagagggggaacctgcccacagcatggggagggccctg 600
Db 541 CTTGTTGCAGATGGCACTTTGTCTGAGAGAGAGGGGAGCTGCCCCAGCATGGGAGGCCCTG 600
QY 601 gaaatccctc-ggattaaactgtgaacactgtgtctcttatacctggaggtgtgtgtgt 660
Db 601 GAAGATCCCTC-GGATTAACTGTGAACACTGTATTGCTGCTTTATACCTGGAGTTGTGTGT 660
QY 661 tatctgggtacacatctgctgggtgaaatgagttcatgggctttatttcagtgaggtattta 720
Db 661 TATCTGGGTACACATCTGCTGGGTGAATGAGTTCTATGGGCTTTATTTTCAGTGAGGTATTTA 720
QY 721 cctgagggaaagaagagactggtgtcccaaaagcacagacttttaaatctgtgggtgtgac 780
Db 721 CCTGAGGAGAAAGAGGACTGCTGGTCCCAACAAAGCACAGCTTTTAAATCTGTGGGTGTGAC 780
QY 781 ccattatgactatcataaactgagtgaggtatcaagaataacttttagcaggtgtgtaaaa 840
Db 781 CCATTATGGACTATCATAAATGAGTGAGGTATCAAGAATACTTTAGCAGGTGTGTAATAA 840
QY 841 gatttttgaa-gcgcaacgacacaaactgaactcaaaaatcaagcatggcgatgcctg 900
Db 841 GATTTTGAATGCGCAACGACCAAACTGAACCTCAAAAATCAAGCATGGCATGGATCCTG 900
QY 901 ggtgctcctgaaagcaacttgcttacttgatgcatgtgagattgacggtagccttggtct 960
Db 901 GGTGCTCCTGTGAAGCACTTGCTCTTACTGCAATGTGGGACTTTCAGCGGTAGCTTGGTCT 960
QY 961 gaaatgcaaaacactggctgttggtgtgacagggccacagccacacgctgctgaaacacct 1020
Db 961 GAATGCACAAACCTGGCTTTGGGCTGACAGGCCACACGCCGCTGCTGAAACACCTC 1020
QY 1021 agctcaggtttgtggctatgtccctatgacttgagcttgagcttactttattgcacatatataat 1080
Db 1021 AGCTCAGGTTTGTGGCTATGCTTATGCTTATGACTTGGACTTACTTTTATTGCACATATAATAT 1080
QY 1081 tttcctgc 1088
Db 1081 TTTCTCTC 1088

```

RESULT 2

AC004021/c

LOCUS

DEFINITION

AC004021 171370 bp DNA

PRI

04-FEB-2000

Human PAC clone RPI-186K10 from 5q31, complete sequence.



```

AC004021
VERSION AC004021.1 GI:2795824
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL
REFERENCE 1 (bases 1 to 171370)
AUTHORS Kalicki, J., Kellen, J. and O'Brien, D.
TITLE The sequence of H. sapiens PAC clone RP1-186K10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 171370)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1998) Department of Genetics, Washington
REFERENCE 3 (bases 1 to 171370)
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington
REFERENCE 4 (bases 1 to 171370)
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington
REFERENCE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE
JOURNAL
COMMENT
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.edu
-----
Summary Statistics
-----
Center project name: HJD0186K10
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
This clone was provided for sequencing by Dr. Michelle M. Le Beau
in the Department of Medicine, University of Chicago, Chicago IL;
and Dr. John D. McPherson, Department of Genetics, Washington
University, St. Louis MO. For additional information about the map
position of this clone, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPI-1, prepared by
Pieter de Jong and coworkers at Roswell Park Cancer Institute,
using the method described by Ioannou et al., Nature Genetics
6:84-9 (1994). The library is from one male donor. For further
details, see http://pacpac.med.buffalo.edu/
The clone is available from Genome Systems, Inc.
(http://www.genomesystems.com).
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of RP1-186K10;
actual end is at 171370 of RP1-186K10. The orientation of this
clone is unknown.

This clone contains STS AFM350yb1 (NID:g1051891) and EST53547
(NID:g1347444).
Location/Qualifiers
1. .171370
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/chromosome="5"
/clone="RPI-186K10"
/clone_lib="RPI-1"
/map="5q31"
1. .1157
/rpt_family="L1"
repeat_region
1290. .2612
/rpt_family="L1"
repeat_region
2617. .2679
/rpt_family="MER81"
repeat_region
3511. .3643
/rpt_family="MER1_type"
repeat_region
3791. .4239
/rpt_family="L2"
repeat_region
4249. .4551
/rpt_family="MaLR"
repeat_region
4553. .4936
/rpt_family="Retroviral"
repeat_region
5017. .5328
/rpt_family="MER2_type"
repeat_region
5361. .5408
/rpt_family="Retroviral"
repeat_region
5409. .5542
/rpt_family="MaLR"
repeat_region
5546. .5893
/rpt_family="L2"
repeat_region
5920. .5994
/rpt_family="MaLR"
repeat_region
6030. .6172
/rpt_family="L2"
repeat_region
6193. .6256
/rpt_family="L2"
repeat_region
6299. .6504
/rpt_family="L2"
repeat_region
6505. .6536
/rpt_family="(TGAA)n"
repeat_region
7620. .8444
/rpt_family="L1"
repeat_region
8919. .8998
/rpt_family="Retroviral"
repeat_region
9041. .10574
/notes="Cpg island (%GC=73.7, o/e=0.70, #CpGs=115)"
10335. .10420
/rpt_family="Alu"
repeat_region
10416. .10523
/rpt_family="Alu"
repeat_region
10535. .10594
/rpt_family="Alu"
repeat_region
10732. .10988
/rpt_family="L1"
repeat_region
11023. .14892
/rpt_family="L1"
repeat_region
15585. .15760
/rpt_family="MER1_type"
repeat_region
18900. .18902
/rpt_family="MIR"
repeat_region
18903. .18949
/rpt_family="L2"
repeat_region
18950. .18972
/rpt_family="MIR"
repeat_region
19618. .19768
/rpt_family="L1"
repeat_region
19769. .20895
/rpt_family="L1"
repeat_region
20863. .22033
/rpt_family="L1"
repeat_region
22147. .24390
/rpt_family="L1"
repeat_region
24614. .25397
/rpt_family="L1"
repeat_region
25396. .26577
/rpt_family="L1"

```

```

repeat_region 27204..27824
/rpt_family="L1"
repeat_region 28298..28456
/rpt_family="Alu"
repeat_region 28460..28510
/rpt_family="7SLRNA"
repeat_region 28548..28834
/rpt_family="Alu"
repeat_region 28855..29167
/rpt_family="Alu"
repeat_region 29768..30055
/rpt_family="L1"
repeat_region 31184..31477
/rpt_family="Alu"
repeat_region 32124..32489
/rpt_family="MaLR"
repeat_region 32647..32733
/rpt_family="MIR"
repeat_region 33532..33594
/rpt_family="L2"
repeat_region 33613..33706
/rpt_family="L1"
repeat_region 34315..34609
/rpt_family="Alu"
repeat_region 34782..34932
/rpt_family="L2"
repeat_region 34935..35152
/rpt_family="MIR"
repeat_region 36512..36802
/rpt_family="Alu"
repeat_region 36882..37126
/rpt_family="L2"
repeat_region 37957..38046
/rpt_family="L2"
repeat_region 38742..38863
/rpt_family="MIR"
repeat_region 40308..40599
/rpt_family="Alu"
repeat_region 41127..41200
/rpt_family="MaLR"
repeat_region 41356..41728
/rpt_family="MaLR"
repeat_region 41917..42152
/rpt_family="L2"
repeat_region 42169..42459
/rpt_family="Alu"
repeat_region 42468..42623
/rpt_family="L2"
repeat_region 43252..43548
/rpt_family="Alu"
repeat_region 43535..43594
/rpt_family="(TAA)n"

Query Match 13.7%; Score 148.6; DB 9; Length 171370;
Best Local Similarity 66.4%; Pred. No. 6.2e-34;
Matches 273; Conservative 0; Mismatches 109; Indels 29; Gaps 3;

Qy 1 agtggggacacagctgctgattgaagatgtttcttgggaaagagtgtaagcctaag 60
|||||
Db 54904 AGTTGGGGACAGAGCTGCTGTGATTAAGTGTTCTCTGGGAAATGAATTAAGCCTAATG 54845

Qy 61 atttccaaaggaaggaagcctaattgggaggaagcctaattgttgaattggga-cacctgc 119
|||||
Db 54844 ATTTTCAATGAAGGAGGAGCTGCTAATTTGGGAGGAGTGCTGCTTAATTTGGGATCATCTGC 54785

Qy 120 gggtaataaagctctctccagctgctcttctctgttttggctctgggagggcgaagc 179
|||||
Db 54784 AGGTAATTAAGGCTCTCTCTGATGGCCCTTCCCTGCTTATGACTCTGGGAGGTTAAGAT 54725

Qy 180 attgagggatgcagagcattctaag-ggctggttcttgggttctctccctccctctgtc 238
|||||
Db 54724 AATGTGAGTATACAGAGCTTCTCAGGCCCTCTGTGAGAACCTCTCCCTTGGACAGAGGG 54665

```

```

Qy 239 caaactcagtgaggtatccctgtctgtgctgctctagtagtgcgctcct----- 287
|||
Db 54664 CCAAGGCAACCCAACTAGGGTACCTTCTTAGTCTTGGAAATAGTTAGAGTGATGT 54605

Qy 288 -----gaggccttggtgagttaaaggtctctctgagctgagctgctcagg 331
|||
Db 54604 CTTCCATAGACCCCTCAGAAAGCTTGGCGAGTTTAGGCTCTCTGGAGCAAGCTGCTGCCAG 54545

Qy 332 gaaacgcagtgctcattgtgaaagggagagaccaggcaaggtgttggctg 382
|||
Db 54544 GGAGTCTGTGAGCCCTCTGGAAGGGAGAGAAAGGAATGAGGTATCATG 54494

RESULT 3
AF185100 AF185100 405 bp DNA MAM 01-JAN-2000
LOCUS Lasiorhinus latifrons microsatellite Lla25CA sequence.
DEFINITION AF185100
ACCESSION AF185100
VERSION AF185100.1 GI:6651417
KEYWORDS southern hairy-nosed wombat.
SOURCE Lasiorhinus latifrons
ORGANISM Lasiorhinus latifrons
Mammalia; Metatheria; Diprotodontia; Vombatidae; Lasiorhinus.
REFERENCE 1 (bases 1 to 405)
AUTHORS Beheregey,L.B., Sunnucks,P., Alpers,D.L. and Taylor,A.C.
TITLE Microsatellite loci for the hairy-nosed wombats (Lasiorhinus krefftii and Lasiorhinus latifrons)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 405)
AUTHORS Taylor,A.C.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-1999) Biological Sciences, Monash University, Wellington Rd., Clayton, VIC 3168, Australia
FEATURES
source
1..405
/organism="Lasiorhinus latifrons"
/db_xref="taxon:30668"
repeat_region 62..103
/note="microsatellite Lla25CA"
/rpt_type=tandem
/rpt_unit=ca
BASE COUNT 139 a 109 c 65 g 92 t
ORIGIN

Query Match 8.4%; Score 91.8; DB 3; Length 405;
Best Local Similarity 66.5%; Pred. No. 6.9e-17;
Matches 147; Conservative 0; Mismatches 72; Indels 2; Gaps 1;

Qy 788 ggactatcataactgagtgaggtatcaagaatacttagcagggtgtaaaagattttt 847
|||||
Db 112 GGTCTCAAGTAAGTGTGGGGGTAGTGAGAACTTTGGCAACAG--TAAAAGGTTTCT 169

Qy 848 gaatgcgaacgacacaaactgaactcaaaaatcaagcatggcagtgatcctgggtctc 907
|||||
Db 170 GAACACACACAGCCAAAATAATTAATAAATCAATGATATAATAATCAAGATGTTT 229

Qy 908 ctggaagcacctgaccttactgcatgtgagcattgacggtagccttggttctgaatgca 967
|||||
Db 230 CTGGCAGCACCTGGCTGTGTTGCATTGCACAACTTCACTGTAGCTTTGTTCTCAACATA 289

Qy 968 caaacctggycttttggctgcacagggccaccagccgctgc 1008
|||||
Db 290 CAACATTCATCTTTGGCACTGCACATGCCATCACACCATAC 330

RESULT 4
AC020862/c AC020862 200094 bp DNA HTG 29-MAR-2000
LOCUS Mus musculus clone RP23-248G14, WORKING DRAFT SEQUENCE, 52
DEFINITION unordered pieces.
ACCESSION AC020862

```

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

AC020862.2 GI:7340301  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 200094)  
DOE Joint Genome Institute.  
Sequencing of Mouse  
Unpublished  
2 (bases 1 to 200094)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Mar 29, 2000 this sequence version replaced gi:5686446.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----Summary Statistics  
Consensus quality: 131468 bases at least Q40  
Consensus quality: 161121 bases at least Q30  
Estimated insert size: 172522 bases at least Q20  
Estimated insert size: 200094; sum-of-contigs estimation  
Estimated insert size: 194000; pulse field gel estimation  
Quality coverage: 3.14x in Q20 bases; pulse field gel estimation  
Quality coverage: 3.05x in Q20 bases; sum-of-contigs estimation  
-----

NOTE: This is a 'working draft' sequence. It currently  
\* consists of 52 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1123: contig of 1123 bp in length  
1124 2480: gap of unknown length  
2481 3689: gap of unknown length  
3690 4967: gap of unknown length  
4968 5997: gap of unknown length  
5998 7349: gap of unknown length  
7350 8832: gap of unknown length  
8833 9879: gap of unknown length  
9880 11185: gap of unknown length  
11186 12905: gap of unknown length  
12906 14190: gap of unknown length  
14191 15635: gap of unknown length  
15636 16930: gap of unknown length  
16931 18376: gap of unknown length  
18377 20346: gap of unknown length  
20347 22028: gap of unknown length  
22029 23853: gap of unknown length  
23854 25196: gap of unknown length

25197 26712: contig of 1516 bp in length  
26713 28473: gap of unknown length  
28474 29662: gap of unknown length  
29663 31346: contig of 1189 bp in length  
31347 32719: gap of unknown length  
32720 34057: gap of unknown length  
34058 36039: gap of unknown length  
36040 37862: gap of unknown length  
37863 39563: gap of unknown length  
39564 41018: gap of unknown length  
41019 42986: gap of unknown length  
42987 44994: gap of unknown length  
44995 47720: gap of unknown length  
47721 50900: gap of unknown length  
50901 54324: gap of unknown length  
54325 58498: gap of unknown length  
58499 62290: gap of unknown length  
62291 67699: gap of unknown length  
67700 72719: gap of unknown length  
72720 78606: gap of unknown length  
78607 85624: gap of unknown length  
85625 92160: gap of unknown length  
92161 98561: gap of unknown length  
98562 105678: gap of unknown length  
105679 111168: gap of unknown length  
111169 119079: gap of unknown length  
119080 128003: gap of unknown length  
128004 134621: gap of unknown length  
134622 142334: gap of unknown length  
142335 152949: gap of unknown length  
152950 161984: gap of unknown length  
161985 171417: gap of unknown length  
171418 185089: gap of unknown length  
185090 200094: gap of unknown length

FEATURES  
Location/Qualifiers  
1..200094  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-248G14"

BASE COUNT 57012 a 42625 c 42956 g 57350 t 151 others









AC025874 193816 bp DNA HTG 26-MAY-2000  
Mus musculus clone RP23-58C21, WORKING DRAFT SEQUENCE, 21 unordered  
pieces.  
AC025874  
AC025874.3 GI:8077097  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 193816)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Mus musculus, clone RP23-58C21  
Unpublished  
2 (bases 1 to 193816)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,  
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campbell,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galegari,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,  
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczky,J.,  
Levine,R., Lieuc,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Melarim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Pisani,C., Pollara,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zaminou,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7656783.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L8065  
Center clone name: 58\_C\_21  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 181059 bases at least Q40  
Consensus quality: 187592 bases at least Q30  
Consensus quality: 190121 bases at least Q20  
Insert size: 188000; agarose-fp  
Insert size: 191816; sum-of-contigs  
Quality coverage: 4.5 in Q20 bases; agarose-fp  
Quality coverage: 4.4 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1393: contig of 1393 bp in length  
\* 1394 1493: gap of 100 bp



```
misc_feature 53873..63432
/note="assembly_fragment"
misc_feature 53533..78287
/note="assembly_fragment"
misc_feature 78388..94313
/note="assembly_fragment"
misc_feature 94414..110412
/note="assembly_fragment"
misc_feature 110513..129036
/note="assembly_fragment"
misc_feature 129137..144731
/note="assembly_fragment"
misc_feature 144832..166561
/note="assembly_fragment"
/note="assembly_fragment"
vector_end:77
vector_side:left
misc_feature 166662..193816
/note="assembly_fragment"
/note="assembly_fragment"
BASE COUNT 56048 a 39960 c 39376 g 56427 t 2005 others
ORIGIN

Query Match 4.3% Score 46.6; DB 50; Length 193816;
Best Local Similarity 54.0%; Pred. No. 0.0086;
Matches 142; Conservative 0; Mismatches 114; Indels 7; Gaps 2;

QY 748 caaagcacgcttttaaatctgtgggtgtgaccattatggactatcataactgagtc 807
|| || || || || || || || || || || || || || || || || || || || ||
Db 83487 CATATACACCTCTTAAACTGTGGGTATTGACCTCACATGG-----CATAAATGAATCT 83433

QY 808 aggtatcaagaatacttagcagtggtgtaaaagattttgaatgc--gcaacgaccaa 865
|| || || || || || || || || || || || || || || || || || || || ||
Db 83432 GAGTGTGATGAAAAAATAGCAGAGCAAAATGTTCTCAACATGCAACAGCAAAATAA 83373

QY 866 actgaactaaaaatcaagcatggatcctggtgtcctgtggaagcacttgcttt 925
|| || || || || || || || || || || || || || || || || || || || ||
Db 83372 ACTCAAAATGAACACCATTAAGTAGTCAAAATCCTAAGTGTCTTGCGAATACGCTCTAT 83313

QY 926 actcatgtgacactgaggtgagctgtgttcttgatgaatcacacacgctggccttgg 985
|| || || || || || || || || || || || || || || || || || || || ||
Db 83312 GTGTAATATATGACTTATTACACCTTGATTCTGTATCATACACAAATGATGACTTTGCA 83253

QY 986 ctgcacagggccaccacgcctgtc 1008
|| || || || || || || || || || || || || || || || || || || || ||
Db 83252 CCGTGTATGATATTGCTGCTATGC 83230

RESULT 11
AC016464
LOCUS AC016464 207160 bp DNA HTG 01-MAR-2000
DEFINITION Mus musculus chromosome 11 clone RP23-409J21 map 11, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION AC016464
VERSION AC016464.3 GI:7137319
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 207160)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 1, 2000 this sequence version replaced gi:6970326.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3198
Center clone name: 409_J-21
----- Summary Statistics
Sequencing vector: M13; M77615; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 196848 bases at least Q40
Consensus quality: 202204 bases at least Q30
Consensus quality: 204082 bases at least Q20
Insert size: 210000; agarose-ff
Insert size: 205160; sum-of-contigs
Quality coverage: 4.8 in Q20 bases; agarose-ff
Quality coverage: 4.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 541: contig of 541 bp in length
* 542 641: gap of 100 bp
* 642 1859: contig of 1218 bp in length
* 1860 1959: gap of 100 bp
* 1960 4152: contig of 2193 bp in length
* 4153 4252: gap of 100 bp
* 4253 6291: contig of 2039 bp in length
* 6292 6391: gap of 100 bp
* 6392 10537: contig of 4146 bp in length
* 10538 10637: gap of 100 bp
* 10638 14144: contig of 3507 bp in length
* 14145 14244: gap of 100 bp
* 14245 18479: contig of 4235 bp in length
* 18480 18579: gap of 100 bp
* 18580 25538: contig of 6959 bp in length
* 25539 25638: gap of 100 bp
* 25639 31995: contig of 6357 bp in length
* 31996 32095: gap of 100 bp
* 32096 40133: contig of 8038 bp in length
* 40134 40233: gap of 100 bp
* 40234 48264: contig of 8031 bp in length
* 48265 48364: gap of 100 bp
* 48365 55351: contig of 6987 bp in length
* 55352 55451: gap of 100 bp
* 55452 63846: contig of 8395 bp in length
* 63847 63946: gap of 100 bp
* 63947 75456: contig of 11510 bp in length
* 75457 75556: gap of 100 bp
* 75557 88148: contig of 12592 bp in length
* 88149 88248: gap of 100 bp
* 88249 99481: contig of 11233 bp in length
* 99482 99581: gap of 100 bp
* 99582 115973: contig of 16392 bp in length
* * * * *
```



Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 16986 bases at least Q40  
Consensus quality: 181928 bases at least Q50  
Consensus quality: 186531 bases at least Q20  
Insert size: 174000; agarose-fp  
Insert size: 188960; sum-of-contigs  
Quality coverage: 3.6 in Q20 bases; agarose-fp  
Quality coverage: 3.3 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 44 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1335: contig of 1335 bp in length  
\* 1336 1435: gap of 100 bp  
\* 1436 2662: contig of 1227 bp in length  
\* 2663 2762: gap of 100 bp  
\* 2763 3912: contig of 1150 bp in length  
\* 3913 4012: gap of 100 bp  
\* 4013 5088: contig of 1076 bp in length  
\* 5089 5188: gap of 100 bp  
\* 5189 7277: contig of 2089 bp in length  
\* 7278 7377: gap of 100 bp  
\* 7378 9713: contig of 2336 bp in length  
\* 9714 9813: gap of 100 bp  
\* 9814 11721: contig of 1908 bp in length  
\* 11722 11821: gap of 100 bp  
\* 11822 13224: contig of 1403 bp in length  
\* 13225 13324: gap of 100 bp  
\* 13325 15356: contig of 2032 bp in length  
\* 15357 15456: gap of 100 bp  
\* 15457 18147: contig of 2691 bp in length  
\* 18148 18247: gap of 100 bp  
\* 18248 20027: contig of 1780 bp in length  
\* 20028 20127: gap of 100 bp  
\* 20128 21910: contig of 1783 bp in length  
\* 21911 22010: gap of 100 bp  
\* 22011 24571: contig of 2561 bp in length  
\* 24572 24671: gap of 100 bp  
\* 24672 26145: contig of 1474 bp in length  
\* 26146 26245: gap of 100 bp  
\* 26246 28341: contig of 2096 bp in length  
\* 28342 28441: gap of 100 bp  
\* 28442 30470: contig of 2029 bp in length  
\* 30471 30570: gap of 100 bp  
\* 30571 32464: contig of 1894 bp in length  
\* 32465 32564: gap of 100 bp  
\* 32565 35322: contig of 2758 bp in length  
\* 35323 35422: gap of 100 bp  
\* 35423 38971: contig of 3549 bp in length  
\* 38972 39071: gap of 100 bp  
\* 39072 42889: contig of 3798 bp in length  
\* 42890 42969: gap of 100 bp  
\* 42970 46094: contig of 3125 bp in length  
\* 46095 46194: gap of 100 bp  
\* 46195 49641: contig of 3447 bp in length  
\* 49642 49741: gap of 100 bp  
\* 49742 52811: contig of 3070 bp in length  
\* 52812 52911: gap of 100 bp  
\* 52912 56535: contig of 3624 bp in length  
\* 56536 56635: gap of 100 bp  
\* 56636 60021: contig of 3386 bp in length  
\* 60022 60121: gap of 100 bp  
\* 60122 63400: contig of 3279 bp in length  
\* 63401 63500: gap of 100 bp  
\* 63501 68803: contig of 5303 bp in length  
\* 68804 68903: gap of 100 bp  
\* 68904 73775: contig of 4872 bp in length

\* 73776 73875: gap of 100 bp  
\* 73876 78846: contig of 4971 bp in length  
\* 78847 78946: gap of 100 bp  
\* 78847 83679: contig of 4733 bp in length  
\* 83680 83779: gap of 100 bp  
\* 83780 88078: contig of 4299 bp in length  
\* 88079 88178: gap of 100 bp  
\* 88179 91539: contig of 3361 bp in length  
\* 91540 91639: gap of 100 bp  
\* 91640 96912: contig of 5273 bp in length  
\* 96913 97012: gap of 100 bp  
\* 97013 102508: contig of 5496 bp in length  
\* 102509 102608: gap of 100 bp  
\* 102609 107447: contig of 4839 bp in length  
\* 107448 107547: gap of 100 bp  
\* 107548 113975: contig of 6428 bp in length  
\* 113976 114075: gap of 100 bp  
\* 114076 122387: contig of 8312 bp in length  
\* 122388 122487: gap of 100 bp  
\* 122488 127799: contig of 5312 bp in length  
\* 127800 127899: gap of 100 bp  
\* 127900 137277: contig of 9378 bp in length  
\* 137278 137377: gap of 100 bp  
\* 137378 147150: contig of 9773 bp in length  
\* 147151 147250: gap of 100 bp  
\* 147251 156464: contig of 9214 bp in length  
\* 156465 156564: gap of 100 bp  
\* 156565 185843: contig of 9279 bp in length  
\* 185844 165943: gap of 100 bp  
\* 165944 176050: contig of 10107 bp in length  
\* 176051 176150: gap of 100 bp  
\* 176151 193260: contig of 17110 bp in length.

## FEATURES

source

1. 133260  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-747D16"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 1335  
/note="assembly\_fragment"  
1436. 2662  
/note="assembly\_fragment"  
2763. 3912  
/note="assembly\_fragment"  
4013. 5088  
/note="assembly\_fragment"  
5189. 7277  
/note="assembly\_fragment"  
7378. 9713  
/note="assembly\_fragment"  
9814. 11721  
/note="assembly\_fragment"  
11822. 13224  
/note="assembly\_fragment"  
13325. 15356  
/note="assembly\_fragment"  
15457. 18147  
/note="assembly\_fragment"  
18248. 20027  
/note="assembly\_fragment"  
20128. 21910  
/note="assembly\_fragment"  
22011. 24571  
/note="assembly\_fragment"  
24672. 26145  
/note="assembly\_fragment"  
26246. 28341  
/note="assembly\_fragment"  
28442. 30470  
/note="assembly\_fragment"  
30571. 32464  
misc\_feature  
1. 1335  
misc\_feature  
1436. 2662  
misc\_feature  
2763. 3912  
misc\_feature  
4013. 5088  
misc\_feature  
5189. 7277  
misc\_feature  
7378. 9713  
misc\_feature  
9814. 11721  
misc\_feature  
11822. 13224  
misc\_feature  
13325. 15356  
misc\_feature  
15457. 18147  
misc\_feature  
18248. 20027  
misc\_feature  
20128. 21910  
misc\_feature  
22011. 24571  
misc\_feature  
24672. 26145  
misc\_feature  
26246. 28341  
misc\_feature  
28442. 30470  
misc\_feature  
30571. 32464



us-09-101-423a-4.rge

	/clone_lib="RPCI-23 Female Mouse BAC"	
misc_feature	1..1393	
misc_feature	/note="assembly_fragment"	
misc_feature	1494..2783	
misc_feature	/note="assembly_fragment"	
misc_feature	2884..4000	
misc_feature	/note="assembly_fragment"	
misc_feature	4101..7562	
misc_feature	/note="assembly_fragment"	
misc_feature	7663..10960	
misc_feature	/note="assembly_fragment"	
misc_feature	11061..14368	
misc_feature	/note="assembly_fragment"	
misc_feature	14469..18390	
misc_feature	/note="assembly_fragment"	
misc_feature	18491..20832	
misc_feature	/note="assembly_fragment"	
misc_feature	vector_side:left	
misc_feature	20933..26678	
misc_feature	/note="assembly_fragment"	
misc_feature	26779..32311	
misc_feature	/note="assembly_fragment"	
misc_feature	32412..37387	
misc_feature	/note="assembly_fragment"	
misc_feature	37488..45236	
misc_feature	/note="assembly_fragment"	
misc_feature	45337..53772	
misc_feature	/note="assembly_fragment"	
misc_feature	53873..63432	
misc_feature	/note="assembly_fragment"	
misc_feature	63533..78287	
misc_feature	/note="assembly_fragment"	
misc_feature	78388..94313	
misc_feature	/note="assembly_fragment"	
misc_feature	94414..110412	
misc_feature	/note="assembly_fragment"	
misc_feature	110513..129036	
misc_feature	/note="assembly_fragment"	
misc_feature	129137..144731	
misc_feature	/note="assembly_fragment"	
misc_feature	144832..166561	
misc_feature	/note="assembly_fragment"	
misc_feature	vector_side:right	
misc_feature	166662..193816	
misc_feature	/note="assembly_fragment"	
BASE COUNT	56048 a 39960 c 39376 g 56427 t 2005 others	
ORIGIN		
Query Match	4.1%; Score 44.6; DB 50; Length 193816;	
Best Local Similarity	53.6%; Pred.No.0.035;	
Matches 142; Conservative	0; Mismatches 114; Indels 9; Gaps 2;	
QY 748	caagcacagctttaaatcgtgggttgaccattatggaactcatcaactgatgc 807	
Db 24516	CATATACCACCTTCTTAACGTGGGATTGACCTCACATGG-----CATATTGAATCT 24570	
QY 808	aggtaacaagaatacttttagcaggtggtaaaagatttttgaatgcgaacgacca--- 863	
Db 24571	GAGTGTCATGAATAAAAAAATTAGCAAGAGCAAATGTTCCCTCAACATGCACAAGCAAAAT 24630	
QY 864	aacctgaactcaaaaatacaagcatggcatcgatccctgggtgcctctgggaagcacttgct 923	
Db 24631	AAACTCAAAATGAACACTATTAAATAGTCAAAAATCCTTAAGTGTTCCTGGCAATACTGCTCT 24690	
QY 924	ttaactgcattggcacttcagctgacctgggttcgtgaatgcacaacacgctggcctttg 983	
Db 24691	ATGTGGTAAATATGACCTTTACTACACCTTGAITCTTGATCACAAATATGGAATACTTTG 24750	
QY 984	ggctgcacagggccacacgccctgtgc 1008	









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 01:08:45 ; Search time 3299.68 Seconds  
(without alignments)  
2038.650 Million cell updates/sec

Title: us-09-101-423a-4  
Perfect score: 1088  
Sequence: 1 agggggacacagcttgcgt.....acataaaatatttcctgc 1088

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 7189864 seqs, 3091403243 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*  
44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estom.\*  
68: em\_estov1.\*  
69: em\_estov2.\*  
70: em\_estpl1.\*  
71: em\_estpl2.\*  
72: em\_estpl3.\*  
73: em\_estpl4.\*  
74: em\_estpl5.\*  
75: em\_estro1.\*  
76: em\_estro2.\*  
77: em\_estro3.\*  
78: em\_estro4.\*  
79: em\_estro5.\*  
80: em\_estro6.\*  
81: em\_estro7.\*  
82: em\_estro8.\*  
83: em\_estro9.\*  
84: em\_estro10.\*  
85: em\_estro11.\*  
86: em\_estro12.\*  
87: em\_estro13.\*  
88: gb\_gss1.\*  
89: gb\_gss2.\*  
90: gb\_gss3.\*  
91: gb\_gss4.\*  
92: em\_gss1.\*  
93: em\_gss2.\*  
94: em\_gss3.\*  
95: em\_gss4.\*  
96: gb\_gss5.\*  
97: gb\_gss6.\*  
98: gb\_gss7.\*  
99: gb\_gss8.\*  
100: gb\_gss9.\*  
101: em\_gss5.\*  
102: em\_gss6.\*  
103: em\_gss7.\*  
104: em\_gss8.\*  
105: em\_gss9.\*  
106: em\_gss10.\*  
107: em\_gss11.\*  
108: gb\_gss10.\*  
109: gb\_gss11.\*  
110: em\_gss12.\*  
111: gb\_gss12.\*  
112: gb\_gss13.\*  
113: gb\_gss14.\*  
114: gb\_gss15.\*  
115: gb\_gss16.\*  
116: gb\_gss17.\*

117: gb\_gss18:\*  
 118: gb\_gss19:\*  
 119: em\_gss13:\*  
 120: gb\_gss20:\*  
 121: gb\_gss21:\*  
 122: gb\_gss22:\*  
 123: gb\_gss23:\*  
 124: gb\_gss24:\*  
 125: em\_gss14:\*  
 126: em\_gss15:\*  
 127: em\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
C 1	21	1.9	407	115	AZ101663
C 2	21	1.9	435	90	AQ213601
C 3	21	1.9	615	108	AQ731241
C 4	20	1.8	320	6	AA705131
C 5	20	1.8	325	88	AQ039710
C 6	20	1.8	340	6	AA764810
C 7	20	1.8	395	1	AA009924
C 8	20	1.8	402	20	AW193230
C 9	20	1.8	403	5	AA703346
C 10	20	1.8	420	38	R07777
C 11	20	1.8	453	89	AQ164333
C 12	20	1.8	472	9	A1317728
C 13	20	1.8	477	96	AQ372541
C 14	20	1.8	540	108	AQ685313
C 15	20	1.8	649	108	AQ746413
C 16	19	1.7	187	7	AA927637
C 17	19	1.7	195	90	AQ232532
C 18	19	1.7	200	24	AW815027
C 19	19	1.7	220	4	AA514661
C 20	19	1.7	240	16	AV153659
C 21	19	1.7	275	97	AQ246420
C 22	19	1.7	296	16	AV150873
C 23	19	1.7	305	5	AA662748
C 24	19	1.7	305	15	AV099482
C 25	19	1.7	308	22	AW452408
C 26	19	1.7	355	19	AW086655
C 27	19	1.7	366	37	H62898
C 28	19	1.7	373	5	AA687144
C 29	19	1.7	375	90	AQ216167
C 30	19	1.7	377	90	AQ241736
C 31	19	1.7	388	37	H61577
C 32	19	1.7	395	11	A1615496
C 33	19	1.7	414	6	AA826288
C 34	19	1.7	430	5	AA561219
C 35	19	1.7	432	118	AZ295151
C 36	19	1.7	435	96	AQ374643
C 37	19	1.7	453	11	A1568657
C 38	19	1.7	456	117	A2265893
C 39	19	1.7	457	8	A1139321
C 40	19	1.7	486	22	AW341537
C 41	19	1.7	490	109	AQ800552
C 42	19	1.7	492	24	AW769235
C 43	19	1.7	520	118	B17021
C 44	19	1.7	525	112	AQ891669
C 45	19	1.7	529	33	BE115984

## ALIGNMENTS

RESULT 1  
 AZ101663/c

LOCUS AZ101663 407 bp DNA GSS 09-MAY-2000  
 DEFINITION RPCI-23-452G22.TV RPCI-23 Mus musculus genomic clone RPCI-23-452G22  
 , DNA sequence.  
 ACCESSION AZ101663  
 VERSION AZ101663.1 GI:7754719  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 407)  
 AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret  
 and and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 JOURNAL Other\_GSSs: RPCI-23-452G22.TJ  
 COMMENT Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301.838 0200  
 Fax: 301.838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@jefferson.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
 or from Resea ch Genetics (info@resgen.com). BAC end page:  
 http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 432 row: G column: 22  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers  
 1. 407  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-452G22"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1;  
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBACe3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 106 a 96 c 80 g 125 t  
 ORIGIN  
 Query Match 1.9%; Score 21; DB 115; Length 407;  
 Best Local Similarity 100.0%; Pred. No. 3.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 433 acgaagtttgagactggcctg 453  
 ||||||||||||||||||||  
 Db 257 ACCAAGTTTGAGACGGCCTG 237  
 RESULT 2  
 AQ213601/c  
 LOCUS AQ213601 435 bp DNA GSS 18-SEP-1998  
 DEFINITION HS\_3218\_32\_E07\_MR Cit Approved Human Genomic Sperm Library D Homo  
 sapiens genomic clone Plate-3218 Col-14 Row-J, DNA sequence.  
 ACCESSION AQ213601  
 VERSION AQ213601.1 GI:3624802  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 435)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 99380589  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Sequence Tagged Connector  
 Plate: 3218 row: J column: 14  
 Class: BAC ends  
 High quality sequence stop: 435.  
 Location/Qualifiers  
 1. 435  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-3218 Col-14 Row-J"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in  
 E-Coli DH10B"  
 BASE COUNT 82 a 126 c 88 g 131 t 8 others  
 ORIGIN  
 Query Match 1.9%; Score 21; DB 90; Length 435;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 437 agttgagactgacctggaa 457  
 |||||  
 Db 134 AGTTTGAGACTGGCCTGGAA 114  
 RESULT 3  
 A0731241  
 LOCUS  
 DEFINITION HS\_5517\_B1\_D09\_7FA RPCI-11 Human Male BAC library Homo sapiens  
 genomic clone Plate-1093 Col-17 Row-H, DNA sequence.  
 ACCESSION A0731241  
 VERSION A0731241.1 GI:5502793  
 KEYWORDS GSS  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 615)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 99380589  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieterdejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))

or from Resear h Genetics ([infoeresgen.com](http://infoeresgen.com)). BAC end Web Server:  
<http://www.htsc.washington.edu>  
 Plate: 1093 row: H column: 17  
 Seq primer: T7  
 Class: FAC ends  
 High quality sequence stop: 615.  
 Location/Qualifiers  
 1. 615  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate-1093 Col-17 Row-H"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACE3.6 vector at EcoRI sites"  
 BASE COUNT 157 a 186 c 130 g 138 t 4 others  
 ORIGIN  
 Query Match 1.9%; Score 21; DB 108; Length 615;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 437 agttgagactgacctggaa 457  
 |||||  
 Db 279 AGTTTGAGACTGGCCTGGAA 299  
 RESULT 4  
 A0705131/c  
 LOCUS  
 DEFINITION zj95h07.s1 Soares\_fetal\_liver\_spleen\_INFLS\_S1 Homo sapiens cDNA  
 clone IMAGE:462685 3', mRNA sequence.  
 ACCESSION A0705131  
 VERSION A0705131.1 GI:2715049  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 320)  
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin  
 J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
 White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 WashU-NCI human EST Project  
 Unpublished (1997)  
 Contact:Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available royalty-free through LML; contact the  
 IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 319.  
 Location/Qualifiers  
 1. 320  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3753708"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:462685"  
 /clone\_lib="Soares\_fetal\_liver\_spleen\_INFLS\_S1"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)  
 With a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
 This is a subtracted version of the original Soares fetal



AA009924  
VERSION AA009924.1 GI:1470971  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B., Chisoso, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Lennan, G., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierriy-Meg, J., Trevas, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1099 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 363.  
Location/Qualifiers  
1. 395  
/organism="Homo sapiens"  
/db\_xref="GDB:1329875"  
/db\_xref="taxon:9606"  
/clone="IMAGE:430106"  
/clone\_lib="Soares\_fetal\_liver\_spleen\_LNFLS\_S1"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACGGAAGATTATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 82 a 126 c 81 g 104 t 2 others  
ORIGIN

Query Match 1.8%; Score 20; DB 1; Length 395;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 tctgagctcctcaggaaa 335  
|||||  
DB 134 TCTGAGCTCCTCAGGGAAA 153

RESULT 8  
AW193230/c  
LOCUS AW193230 402 bp mRNA EST 29-NOV-1999  
DEFINITION x172b06.x1 NCI-CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2680211 3' similar to contains Alu repetitive element, mRNA sequence.  
ACCESSION AW193230  
VERSION AW193230.1 GI:6471929  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B., Chisoso, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Lennan, G., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierriy-Meg, J., Trevas, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1099 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 363.  
Location/Qualifiers  
1. 402  
/organism="Homo sapiens"  
/db\_xref="GDB:2680211"  
/clone\_lib="NCI-CGAP\_Pan1"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #:  
11548-013"

BASE COUNT 94 a 126 c 81 g 99 t 2 others  
ORIGIN

Query Match 1.8%; Score 20; DB 20; Length 402;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 aagtttgagactgcctggg 455  
|||||  
DB 338 AAGTTTGAGACTGCCTGGG 319

RESULT 9  
AA703346/c  
LOCUS AA703346 403 bp mRNA EST 24-DEC-1997  
DEFINITION z11e02.s1 Soares fetal liver spleen\_LNFLS\_S1 Homo sapiens CDNA clone IMAGE:449978 3', mRNA sequence.  
ACCESSION AA703346  
VERSION AA703346.1 GI:2713264  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B., Chisoso, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Lennan, G., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierriy-Meg, J., Trevas, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1099 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 363.  
Location/Qualifiers  
1. 403  
/organism="Homo sapiens"

```

/db_xref="GDB:1354235"
/db_xref="taxon:9606"
/clone_image="449978"
/clone_lib="Soares fetal_liver_spleen_infls_s1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AACTGAAGAATAAATAAGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      105 a      84 c      128 g      86 t
ORIGIN
Query Match      1.8%; Score 20; DB 5; Length 403;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 tctgagctgctcaggaaa 335
|||||
Db 260 TCTGAGCTGCTCAGGAAA 241

RESULT 10
LOCUS R07777 420 bp mRNA EST 05-APR-1995
DEFINITION Y15806.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:126899 5' similar to contains MSRI repetitive element ; , mRNA
sequence.
ACCESSION R07777
VERSION R07777.1 GI:759700
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 420)
AUTHORS Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman
M., Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J.,
Rifkin L., Rohlfing T., Soares M., Tan F., Trevaskis E., Waterston
R., Williamson A., Wohldmann P. and Wilson R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert size: 243
High quality sequence stops: 213 Source: IMAGE Consortium, LLNL This
clone is available royalty-free through LLNL ; contact the IMAGE
Consortium (info@image.llnl.gov) for further information.
Insert Length: 243 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 213.
Location/Qualifiers
1. .420
/organism="Homo sapiens"
/db_xref="GDB:479060"
/db_xref="taxon:9606"
/clone_image="126899"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"

FEATURES
source

```

```

/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5', AACTGAAGAATAAATAAGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      78 a      117 c      101 g      117 t
ORIGIN
Query Match      1.8%; Score 20; DB 38; Length 420;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 316 tctgagctgctcaggaaa 335
|||||
Db 138 TCTGAGCTGCTCAGGAAA 157

RESULT 11
LOCUS A0164333 453 bp DNA GSS 16-OCT-1998
DEFINITION HS_2171_A2_G06_MK CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-2171 Col-12 Row-M, DNA sequence.
ACCESSION A0164333
VERSION A0164333.1 GI:3562528
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 453)
AUTHORS Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
Hood L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2171 row: M column: 12
Class: BAC ends
High quality sequence stop: 453.
Location/Qualifiers
1. .453
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="Plate-2171 Col-12 Row-M"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      118 a      127 c      78 g      128 t      2 others
ORIGIN
Query Match      1.8%; Score 20; DB 89; Length 453;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 299 gagttaaggtctctggatct 318
|||||

```



JOURNAL  
MEDLINE  
COMMENT

scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu

Plate: 1107 row: B column: 20  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 540.

FEATURES  
source

1. .540  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-1107 Col-20 Row-B"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 166 a 83 c 118 g 168 t 5 others

Query Match 1.8%; Score 20; DB 108; Length 540;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 429 atagagcaagcttgagactg 448  
|||||  
DB 387 ATAGAGCAAGTTTGAGACTG 406

RESULT 15  
AQ746413/c

LOCUS AQ746413 649 bp DNA GSS 19-JUL-1999  
DEFINITION HS\_2278\_A2\_E12\_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2278 Col-24 Row-I, DNA sequence.

ACCESSION AQ746413  
VERSION AQ746413.1 GI:5533571  
KEYWORDS GSS.

SOURCE  
human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 649)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589

CONTACT: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: http://www.htsc.washington.edu  
Plate: 2278 row: I column: 24  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 649.

FEATURES  
source

1. .649  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-2278 Col-24 Row-I"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in  
E-Coli DH10B"

BASE COUNT 215 a 69 c 136 g 223 t 6 others

## ORIGIN

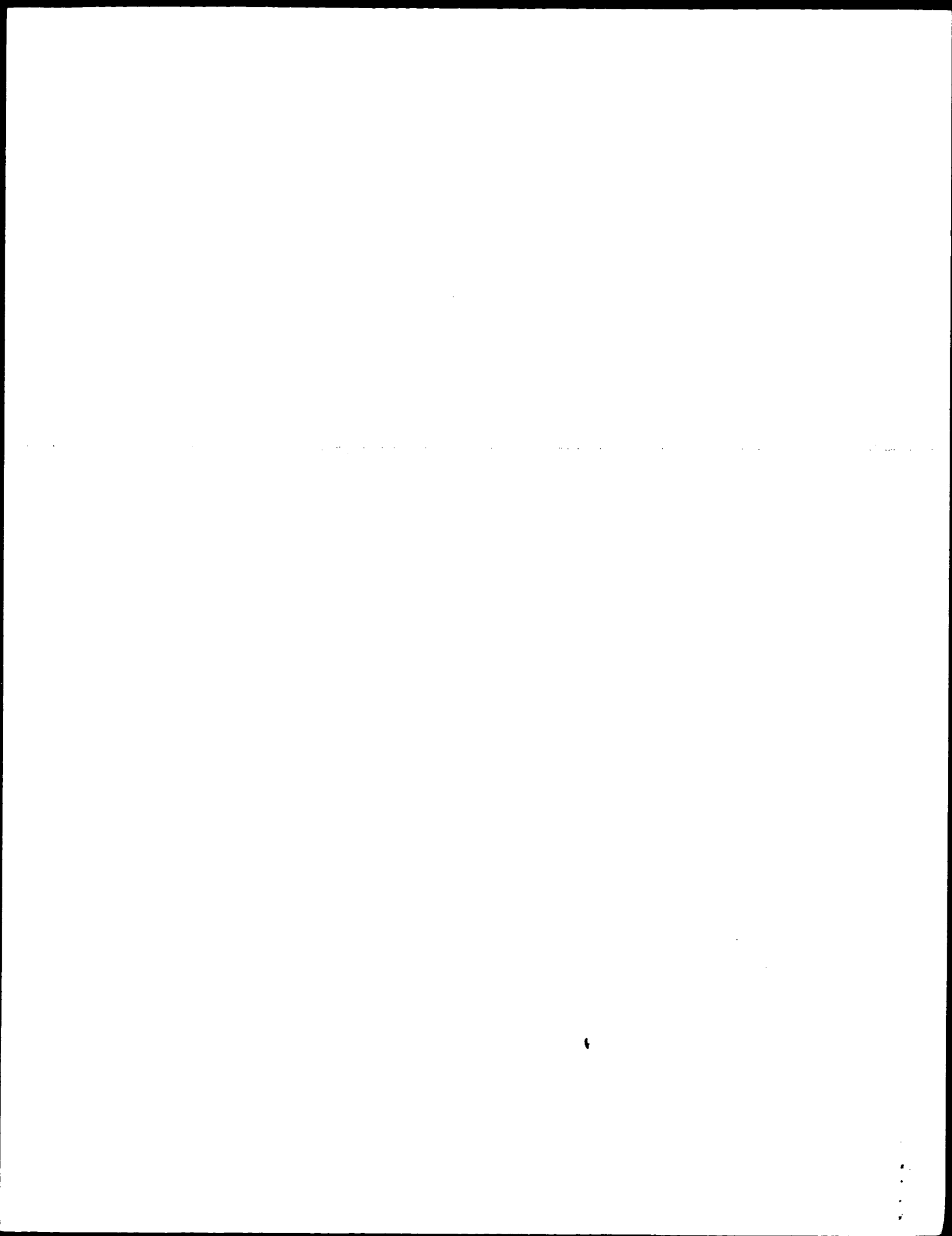
Query Match 1.8%; Score 20; DB 108; Length 649;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 423 ctctattatagacgaagtttg 442  
|||||  
DB 483 CTCATTATAGACGAAGTTTG 464

Search completed: November 4, 2000, 01:08:50  
Job time: 17431 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:38:12 ; Search time 255.06 Seconds  
(without alignments)  
1602.450 Million cell updates/sec

Title: US-09-101-423A-4  
Perfect score: 1088  
Sequence: 1 agtggggacacagcttgct.....acataataattttctgc 1088

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 480022 seqs, 187831343 residues

Word size : 0

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1088	100.0	1088	18	T72788
2	19	1.7	5000	19	Metastasis inducin
3	18	1.7	135	16	Human kynurenine-3
4	18	1.7	675	19	Human gene signatu
5	18	1.7	1014	19	H. pylori cell env
6	18	1.7	1053	18	H. pylori cell env
7	18	1.7	1119	18	H. pylori secreted
8	18	1.7	2992	21	H. pylori secreted
9	18	1.7	3754	18	Human Memopoeitic
10	18	1.7	6971	17	Staphylococcus aur
11	18	1.7	9069	19	Moraxella outer me
12	17	1.6	798	18	Streptococcus pneu
					H. pylori secreted

13	17	1.6	898	19	X14044
c 14	17	1.6	1152	16	O95326
c 15	17	1.6	1183	20	V84547
16	17	1.6	1206	20	V90851
17	17	1.6	1208	20	V90562
c 18	17	1.6	1429	21	A37050
19	17	1.6	1484	18	V74906
20	17	1.6	1753	20	V80184
21	17	1.6	1981	20	V83711
22	17	1.6	2084	16	O88694
23	17	1.6	2143	20	Z42062
c 24	17	1.6	2466	21	Z46438
c 25	17	1.6	2883	21	A37029
26	17	1.6	2923	11	O06173
27	17	1.6	2923	13	O32855
28	17	1.6	2923	14	O37568
29	17	1.6	2923	14	O41294
c 30	17	1.6	3162	16	T02947
c 31	17	1.6	3174	21	A26421
32	17	1.6	3319	19	V63177
c 33	17	1.6	3671	21	A37051
c 34	17	1.6	3906	20	V70207
c 35	17	1.6	4165	17	T18893
c 36	17	1.6	4322	19	V58192
c 37	17	1.6	5631	20	X98274
38	17	1.6	15577	19	V35616
39	17	1.6	19639	20	X23524
40	17	1.6	32367	19	V35620
41	17	1.6	49377	19	V05287
c 42	17	1.6	910715	20	X20248
c 43	16	1.5	310	21	A06642
c 44	16	1.5	337	20	X07455
c 45	15	1.5	352	21	A06640

## ALIGNMENTS

RESULT 1  
T72788  
ID T72788 standard; DNA; 1088 BP.  
XX  
AC T72788;  
XX  
DT 22-SEP-1997 (first entry)  
XX  
DE Metastasis inducing DNA C9.  
XX  
KW Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;  
KW osteopontin; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO9725443-A1.  
XX  
PD 17-JUL-1997.  
XX  
PF 10-JAN-1997; 97WO-GB000074.  
XX  
PR 10-JAN-1996; 96GB-0000470.  
XX  
(UYLI-) UNIV LIVERPOOL.  
XX  
PI Barracough BR, Rudland PS;  
XX  
DR WPI; 1997-372878/34.  
XX  
PT New isolated metastasis-inducing DNA - used to develop products to  
PT identify and treat patients at risk from metastatic tumours  
XX  
PS Claim 11; Page 26; 38pp; English.  
XX  
CC Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12

H. pylori GHPO 141  
Murine GalNAC-4ph  
Human secreted pro  
Nucleotide sequenc  
Human PRO1326 (UNQ  
Staphylococcus aur  
Bone morphogenetic  
Nucleic acid encod  
HindIII fragment p  
Human endometrium  
Human z-dependent  
Human PRO1113 (UNQ  
Human Bone Morphog  
BMP6. Rattus ratt  
Human BMP-6 coding  
Human BMP-6 gene.  
EPH-like receptor  
Human secreted pro  
DNA from clone gm2  
Human PRO1249 (UNQ  
Rat receptor tyros  
Rat REK7 cDNA. Ra  
Mouse Bsk receptor  
Human tumour suppr  
SHOX gene prelinin  
O. longistaminata  
Human SHOX (short  
The soraphen biosy  
Borrelia burgdorfe  
Human immunogenic  
Human BS124 specif  
Human immunogenic

CC and C20 (T72785-90) are entirely novel short stretches of human  
 CC regulatory DNA capable of inducing metastasis. They were  
 CC identified using a method for detecting Met-DNA that involves  
 CC transferring human DNA from a malignant, metastatic cancer cell  
 CC (in this case, breast cancer) into a cell line (pref. rat Rama 37)  
 CC that produces only benign, non-metastasizing tumours when injected  
 CC into a syngeneic animal, infecting the transformed cells into a  
 CC syngeneic animal, selecting those animals having metastasizing  
 CC tumours, and recovering the Met-DNA from them. The isolated  
 CC Met-DNAs can be used to develop products to identify and treat  
 CC patients at risk from metastatic tumours. Hybridisation of C9 DNA  
 CC occurred to HindIII-digested DNA from 4/9 human breast tumours  
 CC examined; there was no hybridisation to HindIII fragments from  
 CC healthy human breast or colon tissue.  
 XX  
 SQ Sequence 1088 BP; 265 A; 223 C; 302 G; 298 T; 0 other;

Query Match 100.0%; Score 1088; DB 18; Length 1088;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1088; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agtggggacacagctgtgctgattgaagatgttcttgggaaagagtgtaagcctaagc 60  
 DB 1 agtggggacacagctgtgctgattgaagatgttcttgggaaagagtgtaagcctaagc 60  
 QY 61 atttccaatgaaagagctgctaatgtgggagcgaatgttcttaattgggacacacctgc 120  
 DB 61 atttccaatgaaagagctgctaatgtgggagcgaatgttcttaattgggacacacctgc 120  
 QY 121 ggttaattaaaagctctctccagtggtcttctctgttttggctctggagcggaagcca 180  
 DB 121 ggttaattaaaagctctctccagtggtcttctctgttttggctctggagcggaagcca 180  
 QY 181 ttgagaggatgacaggaattctcaaggctggtcttctggtttctccctccctctgtcca 240  
 DB 181 ttgagaggatgacaggaattctcaaggctggtcttctggtttctccctccctctgtcca 240  
 QY 241 aactcagtgaggtatccctgtctgtctgctccttagagtcgctctgaggtctggtga 300  
 DB 241 aactcagtgaggtatccctgtctgtctgctccttagagtcgctctgaggtctggtga 300  
 QY 301 gtttaaggtctctgagctgagctgcctcagggaaacgcagctgagctcattggaaagggag 360  
 DB 301 gtttaaggtctctgagctgagctgcctcagggaaacgcagctgagctcattggaaagggag 360  
 QY 361 aacaggcgaaggtgttggctgtgacctcagaattctgagggcgaaggttcaaggctaa 420  
 DB 361 aacaggcgaaggtgttggctgtgacctcagaattctgagggcgaaggttcaaggctaa 420  
 QY 421 ctctcattatagagcaagtttggagactggcctgggaaacaaataataaagtgtgaggt 480  
 DB 421 ctctcattatagagcaagtttggagactggcctgggaaacaaataataaagtgtgaggt 480  
 QY 481 catatgacagcactgaggtgctgtctcctagatagatcataaggacctgctgctgggga 540  
 DB 481 catatgacagcactgaggtgctgtctcctagatagatcataaggacctgctgctgggga 540  
 QY 541 ctgttgcagatggcacttgtgtcagagaggggacctgcccacagatgggagggacctgc 600  
 DB 541 ctgttgcagatggcacttgtgtcagagaggggacctgcccacagatgggagggacctgc 600  
 QY 601 gaagatcctctgatttaactgtgaacactgattgctgttataacctgaggtgtgctgt 660  
 DB 601 gaagatcctctgatttaactgtgaacactgattgctgttataacctgaggtgtgctgt 660  
 QY 661 tatctgtacacatctctgctgtgggtgaatgagttcatggtgtttatttcaagtgtgattta 720  
 DB 661 tatctgtacacatctctgctgtgggtgaatgagttcatggtgtttatttcaagtgtgattta 720  
 QY 721 cctgagagagaaagagctggtgcccacaaagcacagcttttaaatctgtggtgtgac 780  
 DB 721 cctgagagagaaagagctggtgcccacaaagcacagcttttaaatctgtggtgtgac 780

QY 781 ccattatggactatcataaactgagtcaggtatcacaataactttagcaggtgtgtaaaaa 840  
 DB 781 ccattatggactatcataaactgagtcaggtatcacaataactttagcaggtgtgtaaaaa 840  
 QY 841 gattttgaatgcgaacagcaccacaaactgaactcaaaatacaagcatggtgcatgctcgt 900  
 DB 841 gattttgaatgcgaacagcaccacaaactgaactcaaaatacaagcatggtgcatgctcgt 900  
 QY 901 ggtctcctggaagcaacttgccttactgactgcatgtgtggaactgaggtgagcttggttct 960  
 DB 901 ggtctcctggaagcaacttgccttactgactgcatgtgtggaactgaggtgagcttggttct 960  
 QY 961 gaattgcacaacacgtggctgttggctgcacagggccaccgcctgctgaaacacctc 1020  
 DB 961 gaattgcacaacacgtggctgttggctgcacagggccaccgcctgctgaaacacctc 1020  
 QY 1021 agctcaggttggctgtatgctcctatgacttggacttacttatttgcacataataat 1080  
 DB 1021 agctcaggttggctgtatgctcctatgacttggacttacttatttgcacataataat 1080  
 QY 1081 ttctctgc 1388  
 DB 1081 ttctctgc 1388

RESULT 2  
 V20609/c  
 ID V20609 standard; cDNA; 5000 BP.  
 XX V20609;  
 AC AC  
 XX AC  
 DT 19-JUN-1998 (first entry)  
 XX  
 DE Human kynurenine-3-hydroxylase encoding cDNA extended 3'-clone.  
 KW Human; kynurenine-3-hydroxylase; K3OH; kyn-3-OHase; Xenopus laevis;  
 KW oocyte; expression; inhibitor; neurotoxin; quinolinic acid; tryptophan;  
 KW kynurenine acid; anticonvulsant; neuroprotective; ds.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 47..1507  
 FT /\*tag= a  
 FT /product= "kynurenine-3-hydroxylase"  
 XX  
 FN WO9802553-A1.  
 XX  
 PD 22-JAN-1998.  
 XX  
 PF 03-JUL-1997; JWO-EP03589.  
 XX  
 PR 15-JUL-1996; J6GB-0014823.  
 XX  
 PA (PHAA ) PHARMACIA & UPJOHN SPA.  
 PI Benatti L, Cini M, Covini N, Magagnin S, Speciale C;  
 XX  
 DR WPI; 1998-110593/10.  
 DR P-PSDB; W48252.  
 XX  
 PT New nucleic acid encoding kynurenine-3-hydroxylase of rat and human  
 PT - and related vectors, transformed cells and proteins, useful for  
 PT identifying specific inhibitors, potentially useful for preventing  
 PT biosynthesis of the neurotoxin quinolinic acid  
 XX  
 PS Example 4; Page 45-50; 80pp; English.  
 XX  
 CC The present sequence encodes human kynurenine-3-hydroxylase (K3OH). K3OH  
 CC is involved in metabolism of tryptophan to kynurenine acid (KYNA), which  
 CC has anticonvulsant and neuroprotective activities, and quinolinic acid  
 CC (QUIN) which is neurotoxic, and is a target for inhibition of QUIN

CC synthesis. K3OH DNA is used to produce recombinant K3OH (for  
CC characterisation, screening for potential inhibitors and rational drug  
CC design); to isolate similar sequences from other species; to transform  
CC cells that do not normally produce K3OH; to screen for restriction  
CC fragment polymorphisms associated with specific diseases and for  
CC diagnostic assay of K3OH expression in tissues. Inhibition of K3OH  
CC prevents formation of QUIN and simultaneously causes accumulation of  
CC K3NA.

XX Sequence 5000 BP; 1548 A; 996 C; 954 G; 1502 T; 0 other;

Query Match 1.7%; Score 19; DB 19; Length 5000;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 agtttgagactggcctggg 455  
|||||  
DB 2361 AGTTTGAGACTGGCTGGG 2343

RESULT 3  
T26741/c  
ID T26741 standard; cDNA to mRNA; 135 BP.

XX AC T26741;  
XX DT 23-OCT-1996 (first entry)

DE Human gene signature HUMGS08991.

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
XX human; cloning; mapping; non-biased library; diagnosis; detection;  
XX cell typing; abnormal cell function; ss.

XX Homo sapiens.

XX WO9514772-A1.

XX 01-JUN-1995.

XX 11-NOV-1994; 94WO-JP01916.

XX 12-NOV-1993; 93JP-0355504.

XX (MATSU) MATSUBARA K.

XX (OKUBU) OKUBO K.

XX Matsubara K, Okubo K;

XX WPI: 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
XX for diagnosis of abnormal cell function, by preparing cDNA that  
XX reflects relative abundance of corresp. mRNA in specific human  
XX tissues

XX Claim 1; Page 2161-2162; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.  
XX double-stranded DNA) which comprises one of the 7837 "GS" sequences  
XX given in T19001-T26837 and which is able to hybridise to part of  
XX human genomic DNA, cDNA or mRNA is claimed. The GS (gene signature)  
XX sequences were obtained from 3'-directed cDNA libraries prepared  
XX from various human tissues; synthesis of cDNA was initiated from the  
XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
XX untranslated sequence is unique to a particular mRNA species, almost  
XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
XX is constructed so as to reflect accurately the relative abundance of  
XX different mRNAs in the particular tissue from which it was derived.  
XX The appearance frequency of a given GS in a cDNA library can be  
XX determined (esp. using primers and probes derived from the GS  
XX sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.  
XX Sequence 135 BE; 43 A; 23 C; 23 G; 46 T; 0 other;

Query Match 1.7%; Score 18; DB 16; Length 135;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1060 ttttattgcacatatataa 1077  
|||||  
DB 33 TTTTATGCACATATAA 16

RESULT 4  
X30415  
ID X30415 standard; DNA; 675 BP.

XX AC X30415;

XX DT 08-JUN-1999 (first entry)

DE H. pylori cell envelope protein ORF 04apl2016\_25501501\_fl\_1.

XX Vaccine; probe; diagnostic; ORF; cell envelope protein;  
XX secreted protein; cellular protein; ds.

XX Helicobacter pylori.

XX WO9818323-A1.

XX 07-MAY-1998.

XX 28-OCT-1997; 97WO-US19575.

XX 14-JUL-1997; 97US-0891928.

XX 28-OCT-1996; 96US-0739150.

XX 06-DEC-1996; 96US-0759739.

XX (ASTR) ASTRA AB.

XX Alm RA, Smith D;

XX WPI: 1998-271811/24.

XX P-PSDB; Y10948.

XX Helicobacter pylori nucleic acids and proteins - used to develop  
XX products for the detection, prevention and treatment of H. pylori  
XX infections

XX Claims 3, 4; Page 92; 279pp; English.

XX Recombinant or substantially pure preparations of H. pylori polypeptides  
XX are disclosed, together with the nucleic acids encoding them. In all,  
XX 73 ORFs are shown. The proteins are variously cell envelope proteins,  
XX secreted proteins or other cellular proteins. Vaccines containing the  
XX nucleic acids or proteins are claimed, as are probes containing at least  
XX 8 nucleotides from the nucleic acid sequences. The vaccines are useful  
XX for treating or reducing the risk of H. pylori infections, and the  
XX CC Helicobacter in a sample. The products are also of use in screening  
XX for compounds having the ability to interfere with the H. pylori life  
XX cycle or to inhibit H. pylori infection.

XX Sequence 675 BE; 278 A; 87 C; 129 G; 181 T; 0 other;

Query Match 1.7%; Score 18; DB 19; Length 675;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 837 aaagatttttgatgcg 854  
|||||

```

Db 242 aaagattttgaatgcg 259
RESULT 5
X30431
ID X30431 standard; DNA; 1014 BP.
XX
AC X30431;
XX
DT 08-JUN-1999 (first entry)
XX
DE H. pylori cell envelope protein ORF 29ep10720_25501501_c2_33.
XX
KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
secreted protein; cellular protein; ds.
XX
OS Helicobacter pylori.
XX
PN W09818323-A1.
XX
PD 07-MAY-1998.
XX
PF 28-OCT-1997; 97WO-US19575.
XX
PR 14-JUL-1997; 97US-0891928.
XX
PR 28-OCT-1996; 96US-0739150.
XX
PR 06-DEC-1996; 96US-0759739.
XX
PA (ASTR ) ASTRA AB.
XX
PI Alm RA, Smith D;
XX
DR WPI; 1998-271811/24.
XX
DR P-PSDB; Y10964.
XX
XX
PT Helicobacter pylori nucleic acids and proteins - used to develop
products for the detection, prevention and treatment of H. pylori
infections
PT
PS Claims 3, 4; Page 104; 279pp; English.
XX
CC Recombinant or substantially pure preparations of H. pylori polypeptides
are disclosed, together with the nucleic acids encoding them. In all,
73 ORFs are shown. The proteins are variously cell envelope proteins,
secreted proteins or other cellular proteins. Vaccines containing the
nucleic acids or proteins are claimed, as are probes containing at least
8 nucleotides from the nucleic acid sequences. The vaccines are useful
for treating or reducing the risk of H. pylori infections, and the
probes can be used diagnostically for detecting the presence of
Helicobacter in a sample. The products are also of use in screening
for compounds having the ability to interfere with the H. pylori life
cycle or to inhibit H. pylori infection.
XX
SQ Sequence 1014 BP; 418 A; 133 C; 186 G; 277 T; 0 other;

Query Match 1.7%; Score 18; DB 19; Length 1014;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 837 aaagattttgaatgcg 854
Db 242 aaagattttgaatgcg 259
|||||
RESULT 6
T67338
ID T67338 standard; DNA; 1053 BP.
XX
AC T67338;
XX
DT 14-JUL-1997 (first entry)
XX
XX
DE H. pylori secreted or periplasmic protein ORF 11gel0309orf63.
XX

DE H. pylori secreted or periplasmic protein ORF 25501501.aa.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
secretion; periplasmic; ds.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 1..1053
FT /tag= a
FT /note= "no stop codon is given"
XX
PN W09640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
XX
PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR ) ASTRA A3.
XX
PI Berglindh OT, Smith D, Mellgaerd BL;
XX
DR WPI; 1997-052305/05.
XX
DR P-PSDB; W20340.
XX
XX
PT Helicobacter pylori nucleic acid sequences and related
polypeptide(s) - useful for vaccines to treat or prevent H. pylori
infection, and to detect Helicobacter
PT
PS Claim 23; Pages, 226; 1481pp; English.
XX
CC The present sequence encodes a Helicobacter pylori secreted or
periplasmic protein. The protein may be used in a vaccine to prevent or
treat H. pylori infection or to identify H. pylori polypeptide binding
compounds, useful as potential H. pylori life cycle activators or
inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
determined from overlapping contigs generated by mechanically shearing
the bacterial DNA. The sequences were analysed for ORF of at least 180
nucleotides, and the predicted coding regions defined by computer
evaluation. To identify likely H. pylori antigens for vaccine
development, the amino acid sequences predicted from various ORF were
analysed for significant homology to other known or exported membrane
proteins. Having identified and determined the sequences of interest,
particular regions can be isolated from H. pylori by PCR amplification
for recombinant polypeptide production, e.g. in E. coli hosts.
XX
SQ Sequence 1053 BP; 437 A; 134 C; 193 G; 289 T; 0 other;

Query Match 1.7%; Score 18; DB 18; Length 1053;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 837 aaagattttgaatgcg 854
Db 284 aaagattttgaatgcg 301
|||||
RESULT 7
T68086
ID T68086 standard; DNA; 1119 BP.
XX
AC T68086;
XX
DT 16-JUL-1997 (first entry)
XX
DE H. pylori secreted or periplasmic protein ORF 11gel0309orf63.
XX

```

KW Vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
 KW secreted; periplasmic; ds.  
 XX Helicobacter pylori.  
 OS  
 XX  
 XX  
 FH Location/Qualifiers  
 FT 1..1119  
 FT /\*tag= a  
 FT /note= "no stop codon given"  
 XX  
 XX  
 PN W09640893-A1.  
 XX  
 XX 19-DEC-1996.  
 PD  
 XX  
 XX 06-JUN-1996; 96WO-US09122.  
 PF  
 XX  
 XX 01-APR-1996; 96US-0630405.  
 PR  
 XX 07-JUN-1995; 95US-0487032.  
 PR  
 XX (ASTR ) ASTRA AB.  
 PA  
 XX Berglindh OT, Smith D, Mellgaard BL;  
 PI  
 XX WPI: 1997-052306/05.  
 DR  
 XX P-PSDB; W20833.  
 DR  
 XX  
 XX Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter  
 PT  
 XX Claim 23; Pages 891-892; 1481pp; English.  
 PS  
 XX The present sequence encodes a Helicobacter pylori secreted or  
 CC periplasmic protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the  
 CC bacterial DNA. The sequences were analysed for ORF of at least 180  
 CC nucleotides, and the predicted coding regions defined by computer  
 CC evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences  
 CC of interest, particular regions can be isolated from H. pylori by  
 CC PCR amplification for recombinant polypeptide production, e.g. in  
 CC E. coli hosts.  
 XX  
 XX Sequence 1119 BP; 465 A; 139 C; 196 G; 319 T; 0 other;  
 SQ  
 Query Match 1.7%; Score 18; DB 18; Length 1119;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 837 aaagatttttgatgcg 854  
 Db 350 aaagatttttgatgcg 367  
 ||||||||||||||||  
 RESULT 8  
 Z58549  
 ID Z58549 standard; cDNA; 2992 BP.  
 XX  
 AC Z58549;  
 XX  
 XX 05-JUN-2000 (first entry)  
 DT  
 XX Human haemopoietic stem cell regulatory gene SCM3.  
 DE  
 XX

KW SCM3; human; haemopoietic stem cell; HCM; haematopoiesis;  
 KW gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 FH Location/Qualifiers  
 FT 82..1713  
 FT CDS  
 FT /\*tag= a  
 XX  
 XX W0200008145-A2.  
 PN  
 XX 17-FEB-2000.  
 PD  
 XX  
 XX 02-AUG-1999; 99WO-EP05566.  
 PF  
 XX  
 XX 03-AUG-1998; 98US-0128310.  
 PR  
 XX (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 XX  
 XX Cooke MP, Holness CL, Sirenko OI;  
 PI  
 XX WPI: 2000-205705/18.  
 DR  
 XX P-PSDB; Y79140.  
 DR  
 XX Novel genes involved in hematopoietic stem cell regulation useful in  
 PT gene therapy, for increasing the effective dose -  
 PT  
 XX Claim 3b; Fig 4; 81pp; English.  
 PS  
 XX  
 XX The present sequence is that of the cDNA insert in clone SCM3.  
 CC SCM3 was 1 of 3 clones (see also Z58548 and Z58550) identified as  
 CC being preferentially expressed in human haemopoietic stem cells  
 CC (HSC) following expression profiling of a HSC cDNA library. SCM3,  
 CC SCM26 and SCM113 are novel genes involved in HSC regulation, and  
 CC are useful in gene therapy. A claimed method of increasing the  
 CC effective dose of HSC in a mammal involves: obtaining a population  
 CC of CD34+ cells which includes a subpopulation of HSC; introducing  
 CC into the CD34+ cells a polynucleotide (I) encoding a SCM3, SCM26 or  
 CC SCM113 polypeptide (II); obtaining a subpopulation of genetically  
 CC modified stem cells (III) which overexpress (II); and administering  
 CC (III) to a mammal. A claimed method of increasing the effective  
 CC dose of gene modified cells involves: obtaining a population of  
 CC haematopoietic CD34+ cells which includes a subpopulation of HSC;  
 CC introducing (I); introducing a second polynucleotide encoding a  
 CC therapeutic gene; obtaining genetically modified cells capable of  
 CC expressing (II) and the therapeutic gene; and administering the  
 CC genetically modified cells. A claimed method of blocking  
 CC differentiation of mammalian HSC in vitro involves: isolating the  
 CC CD34+ cells; introducing a vector comprising (I); culturing the  
 CC genetically modified CD34+ cells in the presence of a cytokine;  
 CC and selecting cells in which the encoding polypeptide is  
 CC overexpressed whereby differentiation is blocked. Alternatively,  
 CC blocking of differentiation involves: introducing a vector  
 CC comprising (I) into CD34+ cells; allowing expression of (I) in  
 CC the genetically modified cells; and blocking differentiation of  
 CC the cells.  
 XX  
 XX Sequence 2992 BP; 621 A; 922 C; 871 G; 578 T; 0 other;  
 SQ  
 Query Match 1.7%; Score 18; DB 21; Length 2992;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 182 tgagaggatgcaggcat 199  
 Db 2051 tgagaggatgcaggcat 2068  
 ||||||||||||||||  
 RESULT 9  
 V74869/c  
 ID V74869 standard; DNA; 3754 BP.

XX AC V74869;  
 XX AC  
 XX DT 16-MAR-1999 (first entry)  
 XX DT  
 XX DE Staphylococcus aureus contig SEQ ID #558.  
 XX KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 XX OS Staphylococcus aureus.  
 XX AC  
 XX FH Key Location/Qualifiers  
 FT misc\_feature 661..720  
 FT FT /\*tag= a  
 FT FT /note= "these bases represent a line of missing text in  
 FT FT the sequence listing in the specification. They  
 FT FT are included to maintain the nucleotide numbering  
 FT FT given in the specification for this DNA sequence"  
 FT FT 2461..2520  
 FT FT /\*tag= b  
 FT FT /note= "these bases represent a line of missing text in  
 FT FT the sequence listing in the specification. They  
 FT FT are included to maintain the nucleotide numbering  
 FT FT given in the specification for this DNA sequence"  
 XX PN EP786519-A2.  
 XX PD 30-JUL-1997.  
 XX PF 07-JAN-1997; 97EP-0100117.  
 XX PR 05-JAN-1996; 96US-0009861.  
 XX PR (HUMA-) HUMAN GENOME SCI INC.  
 XX PA  
 XX PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 XX PI Rosen CA;  
 XX DR WPI; 1997-374922/35.  
 XX PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 XX PS Claim 1; Page 1496-1498; 3271pp; English.  
 XX CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 XX SQ Sequence 3754 BP; 1124 A; 578 C; 412 G; 1513 T; 127 other;

Query Match 1.7%; Score 18; DB 18; Length 3754;  
 Best Local Similarity 100.0%; Pred. No. 15;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1069 acatataaattttctt 1086  
 DB 2757 ACATATAAATATTTCTT 2740  
 RESULT 10  
 T38740  
 ID T38740 standard; DNA; 6971 BP.  
 XX  
 AC T38740;  
 XX  
 DT 25-JAN-1997 (first entry)  
 XX  
 DE Moraxella outer membrane protein gene.  
 XX KW Outer membrane protein; OMP; immunogen; vaccine; otitis media;  
 KW diagnosis; ss.  
 XX OS Mycobacterium catarrhalis strain 4223.  
 XX FH Key Location/Qualifiers  
 FT CDS 706..6684  
 FT FT /\*tag= a  
 XX PN WO9634960-A1.  
 XX PD 07-NOV-1996.  
 XX PF 29-APR-1996; 36WO-CA00264.  
 XX PR 26-MAR-1996; 36US-0621944.  
 XX PR 01-MAY-1995; 35US-0431718.  
 XX PR 07-JUN-1995; 35US-0478370.  
 XX PA (CONN-) CONNAUGHT LAB LTD.  
 XX PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;  
 XX DR WPI; 1996-506162/50.  
 XX DR P-PSDB; W04505.  
 XX PT Moraxella outer membrane protein - useful as immunogen in protective  
 PT vaccine and for diagnosis  
 XX PS Claim 14; Fig 6; 109pp; English.  
 XX CC A genomic DNA fragment (T38740) of Moraxella catarrhalis otitis  
 CC media strain 4223 includes the coding region for a 1992-amino  
 CC acid protein (W04505) identified as an approx. 200 kDa outer  
 CC membrane protein (OMP). The DNA was isolated from a strain 4223  
 CC genomic library in phage lambda EMBL3 by screening with an  
 CC anti-200 kDa protein guinea pig antiserum. The gene can be  
 CC used for the recombinant expression of the OMP (for use in  
 CC vaccines), for the prepn. of hybridisation probes, or may be  
 CC incorporated into a live vector for use in direct immunisation.  
 XX SQ Sequence 6971 BP; 2265 A; 1553 C; 1533 G; 1620 T; 0 other;  
 Query Match 1.7%; Score 18; DB 17; Length 6971;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 827 gcaggtggtataaagatt 844  
 DB 3823 gcaggtggtataaagatt 3840  
 RESULT 11  
 V52230  
 ID V52230 standard; DNA; 9069 BP.



XX AC V25230;  
 XX AC  
 XX DT 23-OCT-1998 (first entry)  
 XX DE Streptococcus pneumoniae genome fragment SEQ ID NO:97.  
 XX DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KW Streptococcus pneumoniae; S. pneumoniae; vaccine; pharmaceutical composition; ds.  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX OS Streptococcus pneumoniae.  
 XX OS  
 XX PN WO9818931-A2.  
 XX PD 07-MAY-1998.  
 XX PD 30-OCT-1997; 97WO-US19588.  
 XX PF 31-OCT-1996; 96US-0029960.  
 XX PR (HUMA-) HUMAN GENOME SCI INC.  
 XX PA Barash SC, Choi GH, Dillon FU, Dougherty BA, Fannon M;  
 XX PI Kunsch CA, Rosen CA;  
 XX PI WPI; 1998-272225/24.  
 XX DR Computer-readable medium with recorded Streptococcus pneumoniae  
 XX PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 XX PT pharmaceutical compositions and vaccines for Streptococcus  
 XX PT pneumoniae  
 XX PS Claim 1; Page 740-746; 1409pp; English.  
 XX CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1  
 CC to 391, identifying members of the library which contain sequences  
 CC that hybridize to the target sequence and isolating the nucleic acid  
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced  
 CC from an organism, amplifying nucleic acid molecules whose nucleotide  
 CC sequence is homologous to amplification primers derived from the  
 CC fragment of the S. pneumoniae genome to prime the amplification and  
 CC isolating the amplified sequences. The computer readable medium can be  
 CC used in a computer-based system for identifying fragments of the  
 CC S. pneumoniae genome of commercial importance, or expression modulating  
 CC fragments of the S. pneumoniae genome. Products from the present  
 CC invention can be used in diagnosis kits and assays, and pharmaceutical  
 CC compositions and vaccines for S. pneumoniae.  
 XX CC Sequence 9069 BP; 2896 A; 1964 C; 1466 G; 2743 T; 0 other;  
 XX SQ

Query Match 1.7%; Score 18; DB 19; Length 9069;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 tgattgctgtttatacc 646  
 |||||  
 Db 7871 tgattgctgtttatacc 7888

RESULT 12  
 V25078  
 ID V25078 standard; DNA; 798 BP.  
 XX

AC V25078;  
 XX AC  
 XX DT 24-JUN-1998 (first entry)  
 XX DE H. pylori secreted protein ORF hp5p15212\_6928132\_c3\_34.  
 XX DE Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.  
 XX OS Helicobacter pylori.  
 XX OS  
 XX PN Key Location/Qualifiers  
 XX FT 1..798  
 XX FT CDS /\*tag- a  
 XX PD WO9737044-A1.  
 XX PD 09-OCT-1997.  
 XX PD 27-MAR-1997; 97WO-US05223.  
 XX PF 06-DEC-1996; 96US-0761318.  
 XX PR 29-MAR-1996; 96US-0625811.  
 XX PR 02-APR-1996; 96US-0758731.  
 XX PR 25-OCT-1996; 96US-0736905.  
 XX PR 28-OCT-1996; 96US-0738859.  
 XX PA (ASTR ) ASTRA AB.  
 XX PI Alm RA, Smith D;  
 XX PI WPI; 1997-503122/46.  
 XX PR P-PSDB; W55669.  
 XX CC Helicobacter pylori nucleic acid sequences and encoded  
 CC polypeptides) - useful in vaccines to treat or prevent H. pylori  
 CC infection and for diagnosis of H. pylori infection  
 XX CC Claims 5,6,37; Pages 436-437; 1145pp; English.  
 XX CC This sequence encodes a Helicobacter pylori secreted protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions for  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 XX CC Sequence 798 BP; 275 A; 141 C; 174 G; 208 T; 0 other;  
 XX SQ

Query Match 1.6%; Score 17; DB 18; Length 798;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 835 taaaaagattttgaat 851  
 |||||  
 Db 735 taaaaagattttgaat 751

```

RESULT 13
X14044
ID X14044 standard; DNA; 898 BP.
XX
AC X14044;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 1419 gene.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 51..848
FT /*tag= a
XX
PN WO9843478-Al.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
PI WPI; 1998-542293/46.
DR P-PSDB; W98325.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 1; Page 438-439; 2054pp; English.
XX
XX This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 898 BP; 305 A; 151 C; 191 G; 251 T; 0 other;

Query Match 1.6%; Score 17; DB 19; Length 898;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 835 taaaagaatttttgaat 851
DB 785 taaaagaatttttgaat 801

RESULT 14
Q95326/C
ID Q95326 standard; DNA; 1152 BP.
XX
AC Q95326;
XX
DT 11-JUN-1996 (first entry)
XX

```

```

DE Murine GalNac-alpha-2,6-sialyltransferase DNA.
XX
KW GalNac-alpha-2,6-sialyltransferase; anti-inflammatory;
KW viral infection; tumour; migration; inhibitor; sugar chain;
KW reagent; mouse; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 7..1137
FT /*tag= a
XX
PN WO9518217-Al.
XX
PD 06-JUL-1995.
XX
PF 22-DEC-1994; 94WO-JP02182.
XX
PR 28-APR-1994; 94JP-0091507.
PR 24-DEC-1993; 93JP-0348260.
PR 28-MAR-1994; 94JP-0057369.
XX
PA (RIKA) INST PHYSICAL & CHEM RES.
XX
PI Hamamoto T, Kojima N, Kurosawa N, Lee Y, Nakaoka T;
PI Tsuji S;
XX
DR WPI; 1995-246383/32.
DR P-PSDB; R78622.
XX
XX New GalNac-alpha-2,6-sialyltransferases P-B1 and P-B3 - for e.g.
PT treatment of genetic disorders involving missing sugars
XX
XX Example D; Pages 56-58; 70pp; Japanese.
XX
CC Q95326 encodes X78622 murine GalNac-alpha-2,6-sialyltransferase.
CC It can be used as a reagent for introducing human type
CC sugar chains onto proteins, or for the treatment of genetic
CC disorders involving missing sugar chains. It may also be used as
CC a tumour migration inhibitor, viral infection preventative and
CC as an anti-inflammatory.
XX
SQ Sequence 1152 BP; 312 A; 303 C; 281 G; 256 T; 0 other;

Query Match 1.6%; Score 17; DB 16; Length 1152;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 949 agccttggtctgaatg 965
DB 450 AGCCTTGGTTTGAATG 434

RESULT 15
V84547/C
ID V84547 standard; DNA; 1183 BP.
XX
AC V84547;
XX
DT 01-MAR-1999 (first entry)
XX
DE Human secreted protein gene 137 clone HFKEJ07.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX

```

```
OS Homo sapiens.
XX
PN WO9854963-A2.
XX
XX 10-DEC-1998.
XX
XX 04-JUN-1998; 98WO-US11422.
XX 18-DEC-1997; 97US-0070923.
XX 06-JUN-1997; 97US-0048877.
XX 06-JUN-1997; 97US-0048881.
XX 06-JUN-1997; 97US-0048884.
XX 06-JUN-1997; 97US-0048893.
XX 06-JUN-1997; 97US-0048896.
XX 06-JUN-1997; 97US-0048899.
XX 06-JUN-1997; 97US-0048915.
XX 06-JUN-1997; 97US-0048949.
XX 06-JUN-1997; 97US-0048964.
XX 06-JUN-1997; 97US-0048972.
XX 06-JUN-1997; 97US-0048972.
XX 06-JUN-1997; 97US-0049020.
XX 06-JUN-1997; 97US-0049375.
XX 05-SEP-1997; 97US-0057628.
XX 05-SEP-1997; 97US-0057635.
XX 05-SEP-1997; 97US-0057644.
XX 05-SEP-1997; 97US-0057647.
XX 05-SEP-1997; 97US-0057650.
XX 05-SEP-1997; 97US-0057661.
XX 05-SEP-1997; 97US-0057667.
XX 05-SEP-1997; 97US-0057763.
XX 05-SEP-1997; 97US-0057764.
XX 05-SEP-1997; 97US-0057770.
XX 05-SEP-1997; 97US-0057775.
XX 05-SEP-1997; 97US-0057778.
XX 06-JUN-1997; 97US-0048875.
XX 06-JUN-1997; 97US-0048878.
XX 06-JUN-1997; 97US-0048882.
XX 06-JUN-1997; 97US-0048885.
XX 06-JUN-1997; 97US-0048894.
XX 06-JUN-1997; 97US-0048897.
XX 06-JUN-1997; 97US-0048900.
XX 06-JUN-1997; 97US-0048916.
XX 06-JUN-1997; 97US-0048962.
XX 06-JUN-1997; 97US-0048970.
XX 06-JUN-1997; 97US-0048974.
XX 05-SEP-1997; 97US-0049373.
XX 05-SEP-1997; 97US-0057584.
XX 05-SEP-1997; 97US-0057629.
XX 05-SEP-1997; 97US-0057642.
XX 05-SEP-1997; 97US-0057645.
XX 05-SEP-1997; 97US-0057648.
XX 05-SEP-1997; 97US-0057651.
XX 05-SEP-1997; 97US-0057652.
XX 05-SEP-1997; 97US-0057668.
XX 05-SEP-1997; 97US-0057762.
XX 05-SEP-1997; 97US-0057765.
XX 05-SEP-1997; 97US-0057771.
XX 05-SEP-1997; 97US-0057776.
XX 06-JUN-1997; 97US-0048876.
XX 06-JUN-1997; 97US-0048880.
XX 06-JUN-1997; 97US-0048883.
XX 06-JUN-1997; 97US-0048892.
XX 06-JUN-1997; 97US-0048895.
XX 06-JUN-1997; 97US-0048898.
XX 06-JUN-1997; 97US-0048901.
XX 06-JUN-1997; 97US-0048917.
XX 06-JUN-1997; 97US-0048963.
XX 06-JUN-1997; 97US-0048971.
XX 06-JUN-1997; 97US-0049019.
XX 05-SEP-1997; 97US-0049374.
XX 05-SEP-1997; 97US-0057627.
XX 05-SEP-1997; 97US-0057634.
XX 05-SEP-1997; 97US-0057643.
XX 05-SEP-1997; 97US-0057646.

PR 05-SEP-1997; 97US-0057649.
PR 05-SEP-1997; 97US-0057654.
PR 05-SEP-1997; 97US-0057666.
PR 05-SEP-1997; 97US-0057760.
PR 05-SEP-1997; 97US-0057763.
PR 05-SEP-1997; 97US-0057769.
PR 05-SEP-1997; 97US-0057774.
PR 05-SEP-1997; 97US-0057777.
XX
XX (HOMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
XX Fan P, Feng P, Ferrie AM, Fischer CL, Florence C;
XX Florence K, Greene JM, Hu J, Kyaw H, Laflaur DW;
XX Li Y, Moore PE, Ni J, Olsen HS, Rosen CA, Ruben SM;
XX Shi Y, Soppet IR, Wei Y, Young P, Yu G, Zeng Z;
XX
XX WPI; 1999-059865/05.
XX P-PSDB; W88670.
XX
XX New isolated human genes and the secreted polypeptides they encode -
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 4; Page 398; 772pp; English.
XX
XX The invention relates to nucleic acid sequences (V84411 to V84633)
XX encoding human secreted proteins (W88534 to W88756). The secreted protein
XX gene sequences are deposited with the ATCC under deposit numbers ATCC
XX 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
XX 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
XX cells comprising recombinant vectors containing the nucleic acid
XX sequences are used for the recombinant production of the secreted
XX proteins. The polynucleotide and amino acid sequences are useful for are
XX useful for preventing, treating or ameliorating medical conditions e.g.
XX by protein or gene therapy. Pathological conditions can be also diagnosed
XX by determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the polynucleotides, based on
XX which tissues they are most highly expressed in, and include developing
XX products for the diagnosis or treatment of cancer, neurodegenerative
XX disorders, developmental abnormalities and foetal deficiencies, blood
XX disorders, tumours, leukemias, diseases of the immune system, autoimmune
XX diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
XX ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
XX restenosis, prostate diseases, obesity, disorders involving osteoclasts
XX such as osteoporosis, arthritis or malignancies, diseases of testes, lung
XX or thymus, digestive/endocrine disorders, infections and AIDS. The
XX polypeptides are also useful for identifying their binding partners.
XX The present sequence represents a gene encoding a human secreted protein
XX (see descriptor line for gene number and clone identification).
XX
XX Sequence 1183 BP; 221 A; 325 C; 327 G; 302 T; 8 other;
XX
XX Query Match 1.6%; Score 17; DB 20; Length 1183;
XX Best Local Similarity 100.0%; Pred. No. 49;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 365 agcgaagggtgtgtgct 381
XX |||||||
XX Db 479 AGGCAAGGTGTGGCT 463
XX
XX Search completed: November 4, 2000, 03:38:29
XX Job time: 11173 sec
```

us-09-101-423a-4.olig.rng

Mon Nov 6 10:14:25 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:32:45 ; Search time 140.17 Seconds  
(without alignments)  
1173.937 Million cell updates/sec

Title: US-09-101-423A-4  
Perfect score: 1088  
Sequence: 1 agttgggacacagcttgct.....acataataattttctgc 1088

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 262060 seqs, 75620727 residues

Word size : 0

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/PCITUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	ID	Description
C 1	19	1.7	5000	US-09-147-522-5 Sequence 5, Appli
C 2	18	1.7	6973	US-08-478-370-1 Sequence 1, Appli
C 3	17	1.6	1146	US-08-666-367B-4 Sequence 4, Appli
C 4	17	1.6	1659	US-08-943-087-17 Sequence 17, Appli
C 5	17	1.6	2923	US-08-377-292-6 Sequence 6, Appli
C 6	17	1.6	2923	US-07-989-847-7 Sequence 7, Appli
C 7	17	1.6	2923	US-07-989-847-7 Sequence 7, Appli
C 8	17	1.6	3162	US-08-449-645A-12 Sequence 12, Appli
C 9	17	1.6	3162	US-08-702-367A-12 Sequence 12, Appli
C 10	17	1.6	3162	US-08-702-367A-12 Sequence 12, Appli
C 11	17	1.6	3319	US-08-960-022-19 Sequence 19, Appli
C 12	17	1.6	3906	US-08-469-537A-102 Sequence 102, App
C 13	17	1.6	4165	US-08-442-248-1 Sequence 1, Appli
C 14	17	1.6	4165	US-08-440-815-1 Sequence 1, Appli
C 15	17	1.6	4322	US-08-673-789-1 Sequence 1, Appli
C 16	17	1.6	49377	US-08-764-233A-1 Sequence 1, Appli
C 17	16	1.5	186	US-08-090-523-35 Sequence 35, Appli
C 18	16	1.5	186	US-08-398-627-35 Sequence 35, Appli
C 19	16	1.5	522	US-09-130-663-1 Sequence 1, Appli
C 20	16	1.5	1189	US-08-450-042A-2 Sequence 2, Appli
C 21	16	1.5	1371	US-08-687-080-113 Sequence 113, App
C 22	16	1.5	1386	US-08-687-080-113 Sequence 113, App
C 23	16	1.5	1387	US-08-756-506-3 Sequence 3, Appli
C 24	16	1.5	1694	US-08-698-551-9 Sequence 9, Appli
C 25	16	1.5	1694	US-08-602-228-9 Sequence 9, Appli
C 26	16	1.5	1694	US-08-494-440B-9 Sequence 9, Appli

27	16	1.5	1694	3	US-08-533-901B-9	Sequence 9, Appli
28	16	1.5	1694	3	US-08-339-032A-9	Sequence 9, Appli
29	16	1.5	1694	4	US-08-839-031A-9	Sequence 9, Appli
30	16	1.5	1694	6	PCT-US95-12724-9	Sequence 9, Appli
31	16	1.5	1740	4	US-08-471-025-37	Sequence 37, Appli
32	16	1.5	1755	7	5225537-1	Patent No. 5225537
33	16	1.5	2473	2	US-08-698-551-17	Sequence 17, Appli
34	16	1.5	2473	3	US-08-839-032A-17	Sequence 17, Appli
35	16	1.5	2515	3	US-08-695-866-1	Sequence 1, Appli
36	16	1.5	3048	4	US-08-313-200-2	Sequence 2, Appli
37	16	1.5	3048	6	PCT-US93-03837-2	Sequence 2, Appli
38	16	1.5	3307	2	US-08-456-647B-5	Sequence 5, Appli
39	16	1.5	3307	3	US-08-237-401A-5	Sequence 5, Appli
40	16	1.5	3785	2	US-08-445-640-9	Sequence 9, Appli
41	16	1.5	3785	5	US-08-170-558-9	Sequence 9, Appli
42	16	1.5	3785	5	US-08-447-314-9	Sequence 9, Appli
43	16	1.5	3785	5	US-08-445-461-9	Sequence 9, Appli
44	16	1.5	6045	5	US-08-675-566-18	Sequence 18, Appli
45	16	1.5	11725	3	US-08-756-506-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-147-522-5/c  
; Sequence 5, Application US/09147522  
; Patent No. 6107069  
; GENERAL INFORMATION:  
; APPLICANT: MAGAGNIN, SIMONA  
; APPLICANT: BENATTI, LUCA  
; APPLICANT: CINT, MASSIMO  
; APPLICANT: SPECIALE, CARMELA  
; APPLICANT: COVINI, NEVIE  
; TITLE OF INVENTION: RECOMBINANT KYNURENINE-3-HYDROXYLASE ENZYME AND  
; FILE REFERENCE: 0769-0408-0PCT  
; CURRENT APPLICATION NUMBER: US/09/147,522  
; EARLIER FILING DATE: 1999-01-14  
; EARLIER APPLICATION NUMBER: PCT/EP7/03589  
; EARLIER FILING DATE: 1997-07-03  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 5000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (47)..(1507)  
US-09-147-522-5

Query Match 1.7%; Score 19; DB 5; Length 5000;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 agtttgagactggcctggg 455  
|||||  
Db 2361 AGTTTGAGACTGGCCTGGG 2343

RESULT 2  
US-08-478-370-1  
; Sequence 1, Application US/08478370  
; Patent No. 5808024  
; GENERAL INFORMATION:  
; APPLICANT: SASAKI, Ken  
; APPLICANT: HARKNESS, Robin E.  
; APPLICANT: LOCSTORE, Sheena M.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER  
; MEMBRANE PROTEIN OF MORAXELLA

```

;
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,370
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-502
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-370-1
;
; Query Match 1.7%; Score 18; DB 2; Length 6973;
; Best Local Similarity 100.0%; Pred. No. 7.7;
; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 827 gcagtggtgtaaaagatt 844
; | | | | | | | | | | | | | | | |
; DB 3825 GCAGTGGTGAAGATT 3842
;
; RESULT 3
; US-08-666-367B-4/C
; Sequence 4, Application US/08666367B
; Patent No. 5954042
; GENERAL INFORMATION:
; APPLICANT: Shuichi TSUJI et al.
; TITLE OF INVENTION: NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
; TITLE OF INVENTION: PRODUCING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/666,367B
; FILING DATE: August 19, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
;
;
;
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1146
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; ORGANISM: mouse
; FEATURE:
; OTHER INFORMATION: 1-1128 sialyltransferase in soluble
; OTHER INFORMATION: form
; US-08-666-367B-4
;
; Query Match 1.6%; Score 17; DB 3; Length 1146;
; Best Local Similarity 100.0%; Pred. No. 24;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 949 agccttggttctgaatg 965
; | | | | | | | | | | | | | | | |
; DB 444 AGCCITGGTTCTGAATG 428
;
; RESULT 4
; US-08-943-087-17/c
; Sequence 17, Application US/08943087
; Patent No. 5945511
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Kho, Choon J.
; APPLICANT: Jelmsberg, Anna C.
; APPLICANT: Adams, Robyn L.
; APPLICANT: Whitmore, Theodore E.
; APPLICANT: Farrah, Theresa M.
; TITLE OF INVENTION: CYTOKINE RECEPTOR
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,087
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/803,305
; FILING DATE: 20-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: 96-24C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6627
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1659 base pairs
; TYPE: nucleic acid

```

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...1659  
OTHER INFORMATION:  
US-08-943-087-17

Query Match 1.6%; Score 17; DB 4; Length 1659;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 600 ggaagatcctctgatt 616  
|||||  
Db 506 GGAAGATCCTCTGATT 490

RESULT 5  
US-08-377-292-6  
; Sequence 6, Application US/08377292  
; Patent No. 5693615  
; GENERAL INFORMATION:  
; APPLICANT: STONE, ROGER L.  
; TITLE OF INVENTION: THERAPEUTIC FORMULAS FOR OSTEOINDUCTION  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Procter & Gamble Company  
; STREET: 11810 East Miami River Road  
; CITY: Cincinnati  
; STATE: Ohio  
; COUNTRY: U.S.A.  
; ZIP: 45239-8707  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/377,292  
; FILING DATE: 23-JAN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/243,435  
; FILING DATE:  
; APPLICATION NUMBER: US/08/117,367  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Corstange, Brahms J.  
; REGISTRATION NUMBER: 34,804  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 513-245-2858  
; TELEFAX: 513-741-3012  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2923 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-377-292-6

Query Match 1.6%; Score 17; DB 2; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 764 aatctgtgggttgac 780  
|||||  
Db 1102 AATCTGTGGGTGTGAC 1118

RESULT 6  
US-07-989-847-7  
; Sequence 7, Application US/07989847  
; Patent No. 5866364  
; GENERAL INFORMATION:  
; APPLICANT: Israel, David  
; APPLICANT: Wolfman, Neil M.  
; TITLE OF INVENTION: Recombinant Bone Morphogenetic Protein  
; TITLE OF INVENTION: Heterodimers, Compositions and Methods of Use.  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Legal Affairs, Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02140-2387  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Tape  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/989,847  
; FILING DATE:  
; CLASSIFICATION: 436  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rapines, Ellen J.  
; REGISTRATION NUMBER: 32,245  
; REFERENCE/DOCKET NUMBER: GI-5192B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-876-1170  
; TELEFAX: 617-876-5851  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2923 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: Human placenta  
; IMMEDIATE SOURCE:  
; LIBRARY: Stratagene catalog #936203 Human placenta  
; CLONE: BMP6C35  
; POSITION IN GENOME:  
; UNITS: bp  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 160..1701  
; FEATURE:  
; NAME/KEY: met.peptide  
; LOCATION: 1282..1698  
; FEATURE:  
; NAME/KEY: mRNA  
; LOCATION: 1..2923  
US-07-989-847-7

Query Match 1.6%; Score 17; DB 3; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 764 aatctgtgggttgac 780  
|||||  
Db 1102 AATCTGTGGGTGTGAC 1118

RESULT 7  
5187076-5

Patent No. 5187076  
; APPLICANT: WOZNEY, JOHN M.; WANG, ELIZABETH A.; ROSEN, VICKI A.;  
; CELESTE, ANTHONY J.  
; TITLE OF INVENTION: DNA SEQUENCES ENCODING BMP-6 PROTEINS  
; NUMBER OF SEQUENCES: 16  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/490,033  
; FILING DATE: 07-MAR-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 370,544  
; FILING DATE: 23-JUN-1989  
; APPLICATION NUMBER: 347,559  
; FILING DATE: 04-MAY-1989  
; APPLICATION NUMBER: 329,610  
; FILING DATE: 28-MAR-1989  
; APPLICATION NUMBER: 179,100  
; FILING DATE: 08-APR-1988  
; APPLICATION NUMBER: 179,101  
; FILING DATE: 08-APR-1988  
; APPLICATION NUMBER: 179,197  
; FILING DATE: 08-APR-1988  
; APPLICATION NUMBER: 28,285  
; FILING DATE: 20-MAR-1987  
; APPLICATION NUMBER: 31,346  
; FILING DATE: 26-MAR-1987  
; APPLICATION NUMBER: 943,322  
; FILING DATE: 17-DEC-1986  
; APPLICATION NUMBER: 880,776  
; FILING DATE: 01-JUL-1986  
; SEQ ID NO:5  
; LENGTH: 2923  
5187076-5

Query Match 1.6%; Score 17; DB 7; Length 2923;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 764 aatctgtggtgtgtgac 780  
|||||  
Db 1102 aatctgtggtgtgtgac 1118

RESULT 8  
US-08-449-645A-12/c  
; Sequence 12, Application US/08449645A  
; Patent No. 5981245  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Gary M.  
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Patent Operations/RBW  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/449,645A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-287  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 3162 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..2976  
US-08-449-645A-12

Query Match 1.6%; Score 17; DB 4; Length 3162;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 ttgattaagaagtttct 36  
|||||  
Db 2288 TTGATTAAGAAGTTTCT 2272

RESULT 9  
US-08-702-367A-12/c  
; Sequence 12, Application US/08702367A  
; Patent No. 5981246  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Gary M.  
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Patent Operations/RBW  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/702,367A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-287  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3162 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..2976  
US-08-702-367A-12

Query Match 1.6%; Score 17; DB 4; Length 3162;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 ttgattaagaagtttct 36  
|||||  
Db 2288 TTGATTAAGAAGTTTCT 2272

RESULT 10  
PCT-US95-04681-12/c  
; Sequence 12, Application PC/TUS9504681



```

;
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine
; TITLE OF INVENTION: Kinases
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Angen Patent Operations/RBW
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04681
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3162 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2976
; PCT-US95-04681-12

Query Match 1.6% Score 17; DB 6; Length 3162;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 ttgattaagatgtttct 36
|||||
DB 2288 ttgattaagatgtttct 2272

RESULT 11
US-08-960-022-19
; Sequence 19, Application US/08960022
; Patent No. 5976837
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,022
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-960-022-19

Query Match 1.6% Score 17; DB 4; Length 3319;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 cctgtctgtgtctgtcc 272
|||||
DB 2589 CCCTGCTGCTGCTGCC 2605

RESULT 12
US-08-469-537A-102/c
; Sequence 102, Application US/08469537A
; Patent No. 5843745
; GENERAL INFORMATION:
; APPLICANT: Maisonnier, et al.
; TITLE OF INVENTION: EHK AND ROR TYROSINE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill River Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,537A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 08/406,247
; FILING DATE: 17-MAR-1995
; APPLICATION NUMBER: USSN 08/144,992
; FILING DATE: 28-OCT-1993
; APPLICATION NUMBER: USSN 07/736,559
; FILING DATE: 26-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kempler, Ph.D., Gail M
; REGISTRATION NUMBER: 32,143
; REFERENCE/DOCKET NUMBER: REG 070C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3906 base pairs
```

;; TYPE: nucleic acid  
;; STRANDEDNESS: unknown  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: DNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 476-3493  
US-08-469-537A-102

Query Match 1.6%; Score 17; DB 3; Length 3906;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 ttgattaagatgtttct 36  
|||||  
Db 2907 TTGATTAAGATGTTTCT 2891

RESULT 13  
US-08-442-248-1/c  
; Sequence 1, Application US/08442248  
; Patent No. 5759863  
; GENERAL INFORMATION:  
; APPLICANT: Caras, Ingrid W.  
; APPLICANT: Winslow, John W.  
; TITLE OF INVENTION: AL-1 Neurotrophic Factor  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/442,248  
; FILING DATE: 15-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/330128  
; FILING DATE: 27-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, Timothy E.  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: 920C4  
; TELEPHONE: 415/225-8674  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4165 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-442-248-1

Query Match 1.6%; Score 17; DB 2; Length 4165;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 ttgattaagatgtttct 36  
|||||  
Db 2639 TTGATTAAGATGTTTCT 2623

RESULT 14  
US-08-440-815-1/c  
; Sequence 1, Application US/08440815  
; Patent No. 5798448  
; GENERAL INFORMATION:  
; APPLICANT: Caras, Ingrid W.  
; APPLICANT: Winslow, John W.  
; TITLE OF INVENTION: AL-1 Neurotrophic Factor  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/440,815  
; FILING DATE: 15-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/330128  
; FILING DATE: 27-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, Timothy E.  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: 920C1  
; TELEPHONE: 415/225-8674  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4165 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-440-815-1

Query Match 1.6%; Score 17; DB 2; Length 4165;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 ttgattaagatgtttct 36  
|||||  
Db 2639 TTGATTAAGATGTTTCT 2623

RESULT 15  
US-08-673-789-1/c  
; Sequence 1, Application US/08673789  
; Patent No. 5814479  
; GENERAL INFORMATION:  
; APPLICANT: ZHOJ, RENPING; SCHULZ, NICHOLAS,  
; APPLICANT: T.; KROMER, LAWRENCE, F.; VANDE WOUDE,  
; APPLICANT: GEORGE, F.  
; TITLE OF INVENTION: BSK RECEPTOR LIKE  
; TITLE OF INVENTION: TYROSINE KINASE AND LIGAND AND THEIR  
; TITLE OF INVENTION: USE IN DIAGNOSTIC AND THERAPEUTIC  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA

us-09-101-423a-4.olig.rni

Mon Nov 6 10:14:26 2000

```

ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/673,789
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/177,812
FILING DATE: 04-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: CAROL M. GRUPPI
REGISTRATION NUMBER: 37,341
REFERENCE/DOCKET NUMBER: 2026-4105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4322
TYPE: NUCLEOTIDE
STRANDEDNESS: DOUBLE
TOPOLOGY: UNKNOWN
US-08-673-789-1

```

```

Query Match 1.6%; Score 17; DB 2; Length 4322;
Best Local Similarity 100.0%; Pred. NO. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

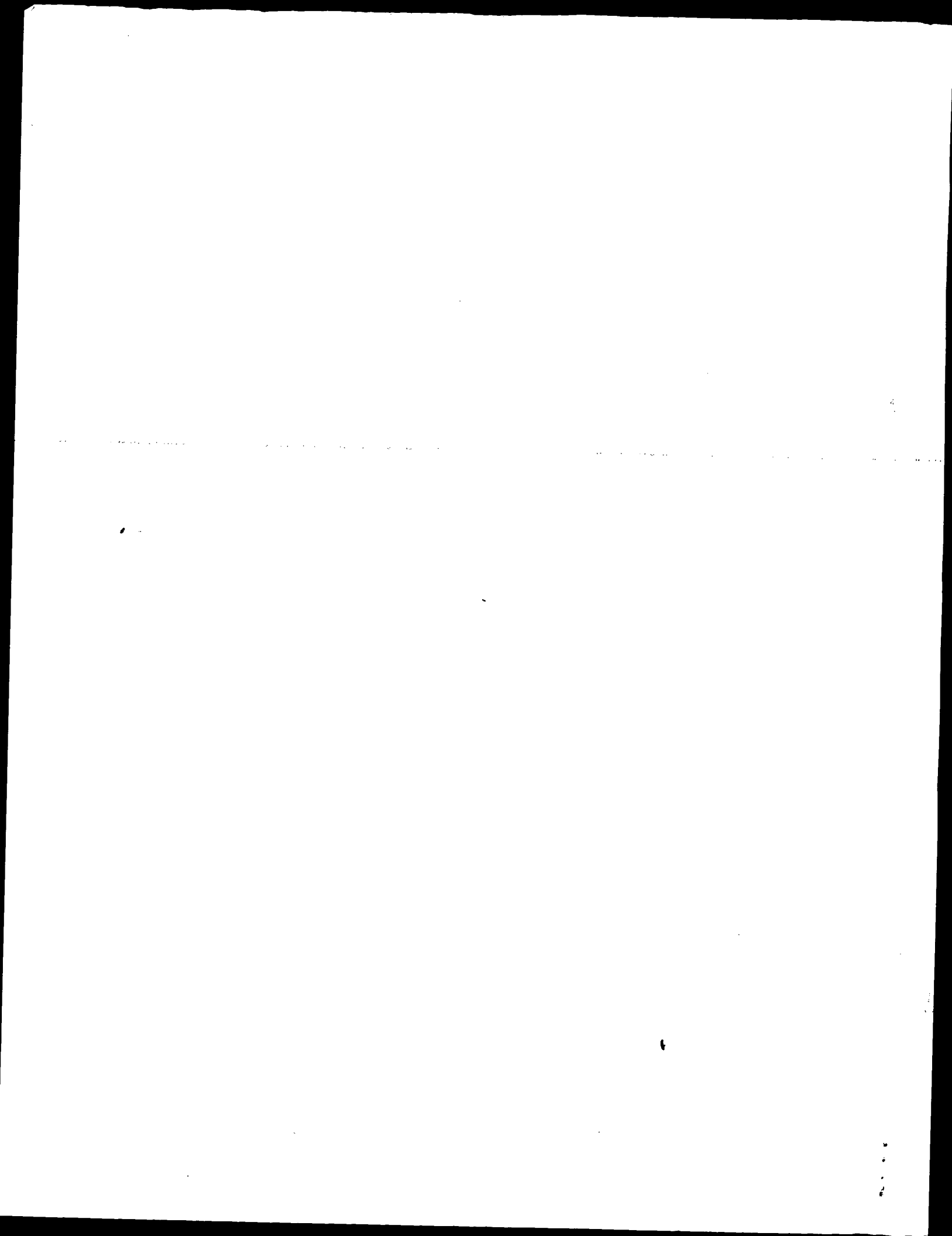
```

```

Oy 20 ttgattaagatgtttct 36
Db 2360 TTGATTAGATGTTTCT 2344

```

Search completed: November 4, 2000, 03:33:06  
Job time: 11976 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2000, 22:08:10 ; Search time 7111.01 Seconds  
(without alignments)  
634.468 Million cell updates/sec

Title: us-09-101-423a-1  
Perfect score: 1033  
Sequence: 1 cttccttggtgctctatgctc.....gtagacataactcccatggt 1033

Scoring table: OLIGO\_NVC  
Gapop 60.0 , Gapext 60.0

Searched: 1033670 seqs, 2183789903 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_bal.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pl3.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: em\_fun.\*  
13: em\_hum1.\*  
14: em\_hum2.\*  
15: em\_in.\*  
16: em\_om.\*  
17: em\_or.\*  
18: em\_ov.\*  
19: em\_pat.\*  
20: em\_ph.\*  
21: em\_pl.\*  
22: em\_ro.\*  
23: em\_sts.\*  
24: em\_sy.\*  
25: em\_un.\*  
26: em\_vi.\*  
27: gb\_htg1.\*  
28: gb\_htg2.\*  
29: gb\_hni.\*  
30: gb\_in2.\*  
31: em\_bal.\*  
32: em\_ba2.\*  
33: em\_hum3.\*  
34: em\_hum4.\*  
35: gb\_pr4.\*  
36: gb\_htg3.\*  
37: gb\_htg4.\*  
38: gb\_htg5.\*  
39: gb\_htg6.\*  
40: gb\_htg7.\*  
41: em\_htg1.\*  
42: em\_htg2.\*  
43: em\_htg3.\*

44: em\_hum5.\*  
45: gb\_pl3.\*  
46: gb\_pr5.\*  
47: gb\_htg8.\*  
48: gb\_htg9.\*  
49: gb\_htg10.\*  
50: gb\_htg11.\*  
51: gb\_htg12.\*  
52: gb\_htg13.\*  
53: gb\_htg14.\*  
54: gb\_in3.\*  
55: gb\_htg15.\*  
56: gb\_htg16.\*  
57: gb\_htg17.\*  
58: em\_htg4.\*  
59: em\_htg5.\*  
60: em\_htg6.\*  
61: em\_htg7.\*  
62: em\_hum6.\*  
63: gb\_htg18.\*  
64: gb\_htg19.\*  
65: gb\_ba3.\*  
66: em\_htg8.\*  
67: em\_htg9.\*  
68: em\_htg10.\*  
69: em\_htg11.\*  
70: em\_htg12.\*  
71: em\_htg13.\*  
72: em\_htg14.\*  
73: em\_htg15.\*  
74: em\_htg16.\*  
75: em\_htg17.\*  
76: em\_htg18.\*  
77: em\_htg19.\*  
78: em\_htg20.\*  
79: em\_htg21.\*  
80: em\_htg22.\*  
81: em\_htg23.\*  
82: gb\_pr6.\*  
83: gb\_pr7.\*  
84: gb\_htg20.\*  
85: gb\_htg21.\*  
86: gb\_htg22.\*  
87: gb\_htg23.\*  
88: gb\_ro.\*  
89: gb\_sts1.\*  
90: gb\_sts2.\*  
91: gb\_sy.\*  
92: gb\_un.\*  
93: gb\_vii.\*  
94: gb\_vii2.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	982	95.1	1033	5	A64058	A64058 Sequence 1
C 2	23	2.2	75335	46	HS107715	AL034394 Human DNA
C 3	21	2.0	2843	82	HSCOLXIX2	U09280 Human type
C 4	21	2.0	51877	11	AL136445	AL136445 Human DNA
C 5	21	2.0	98954	46	HS657J8	AL034407 Human DNA
C 6	21	2.0	120309	88	AC006945	AC006945 Mus muscu
C 7	21	2.0	157684	88	AC004093	AC004093 Mus muscu
C 8	21	2.0	180978	55	AC067861	AC067861 Homo sapi
C 9	21	2.0	193998	55	AC067720	AC067720 Homo sapi
C 10	20	1.9	1447	4	XL095097	U95097 Xenopus lae
C 11	20	1.9	38754	9	AC005173	AC005173 Homo sapi
C 12	20	1.9	99509	54	LMFP1295	AL359773 Leishmani



Mon Nov 6 10:14:00 2000

```
VERSION      AL034394.2  GI:4481839
KEYWORDS     HTG; butyrophillin; CpG Island.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Submitted (08-MAR-1999) Sanger Centre, Hinxton, Cambridgeshire,
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL      1 (bases 1 to 75335)
              Phillips S.
COMMENT      Direct Submission
              CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk
              requests: clonerequest@sanger.ac.uk
              On Mar 23, 1999 this sequence version replaced gi:4090207.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              IMPORTANT: This sequence is not the entire insert of clone 107715.
              It may be shorter because we only sequence overlapping sections
              once, or longer because we arrange for a small overlap between
              neighbouring submissions.
              The true right end of clone dJ17K2 is at 100 in this sequence.
              This sequence has been finished according to sequence map criteria
              as follows. An attempt is made to resolve all sequencing problems,
              such as compressions and repeats, but not necessarily within known
              annotated human repeat sequence elements (e.g. Alu). Where the
              sequence is ambiguous, there is an annotation using the 'unsure'
              feature key.
              This sequence was generated from part of bacterial clone contigs of
              human chromosome 6, constructed by the Sanger Centre Chromosome 6
              Mapping Group. Further information can be found at
              http://www.sanger.ac.uk/HGP/Chr6
              107715 is from the library RPC15 constructed at the Roswell Park
              Cancer Institute by the group of Pieter de Jong. For further
              details see http://daccpac.med.buffalo.edu/VECTOR: pcypac2.

FEATURES             Location/Qualifiers
     source           1..75335
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="p21.2-21.32"
                     /clone="RP5-107715"
                     /clone.lib="RPC1-5"
     repeat_region    740..814
                     /note="5 copies 15 mer 75% conserved"
     gene             join(829..1110,7644..7991,9005..9294)
                     /gene="dJ107715.1"
     CDS              join(<829..1110,7644..7991,9005..9294)
                     /gene="dJ107715.1"
                     /note="match: p78408"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="dJ107715.1 (BTL-II(MHC Class II associated
                     butyrophillin-like))"
                     /protein_id="CAB38473.1"
                     /db_xref="GI:4481840"
                     /db_xref="SPTREMBL:O38261"
                     /translation="LGSAPSIHMEGSGSVQLVCTARGWPEPQVWEDIRGEKLLA
                     VSEHRIQDKGLFYAEATLVVNRNAQSEVSLVNPVLTEKSGVISLPASLKVNGPS
                     QPILVRVGEIDGLTCLSPKANAQSEVYWRDHRYPVAVHYMDGHDVAGQMAEYRG
                     RVLVYSDAIDEGLRFLQILSARPSDDGVRGLFKFDVDVYQASLDLKLVLGSSPLIT
                     VSGDEGEQPMCSDDGVEPQPHVPRDMEGKTIPSSSQALTQSGHGLHFVQTLRLVT
                     NISADVTCSTISIPLGCEKATYFSLGN"
     repeat_region    2801..3100
                     /note="AluJo repeat: matches 2..301 of consensus"
     repeat_region    3102..3403
                     /note="AluJo repeat: matches 1..298 of consensus"
     repeat_region    4266..4474
                     /note="MIR repeat: matches 13..234 of consensus"
     repeat_region    4835..5094
                     /note="L1PA10 repeat: matches 5888..6154 of consensus"

repeat_region       6477..6790
                    /note="MER57B repeat: matches 106..434 of consensus"
repeat_region       7036..7331
                    /note="AluX repeat: matches 1..302 of consensus"
misc_feature        8306..8761
                    /note="match: GSS AQ121363"
repeat_region       11491..11802
                    /note="AluX repeat: matches 1..312 of consensus"
prim_transcript     <12059..12481
                    /note="match: 3' EST AI024674 clone IMAGE:1641755"
repeat_region       12489..12857
                    /note="L1MB5 repeat: matches 5813..6175 of consensus"
repeat_region       12866..13105
                    /note="L1MC2 repeat: matches 6085..6328 of consensus"
prim_transcript     complement(13221..13351)
                    /note="match: 5' EST N51137 clone 281347"
repeat_region       13740..13861
                    /note="L2 repeat: matches 2620..2750 of consensus"
repeat_region       13876..13983
                    /note="L2 repeat: matches 2623..2733 of consensus"
repeat_region       14672..14728
                    /note="MIR repeat: matches 29..89 of consensus"
repeat_region       14873..15008
                    /note="AluSc repeat: matches 1..133 of consensus"
repeat_region       15010..15070
                    /note="L1PA4 repeat: matches 5806..5866 of consensus"
repeat_region       15071..15282
                    /note="L1PA6 repeat: matches 5927..6138 of consensus"
repeat_region       15287..15473
                    /note="AluSc repeat: matches 119..307 of consensus"
repeat_region       17473..17678
                    /note="MIR repeat: matches 12..224 of consensus"
repeat_region       17894..18815
                    /note="SVA repeat: matches 521..1386 of consensus"
misc_feature        18050..19257
                    /note="Putative CpG island"
repeat_region       18833..19826
                    /note="SVA repeat: matches 3..954 of consensus"
repeat_region       20753..21245
                    /note="MLT1F repeat: matches 35..541 of consensus"
repeat_region       21378..21687
                    /note="AluJb repeat: matches 6..295 of consensus"
repeat_region       22203..23003
                    /note="L1MB8 repeat: matches 5255..6164 of consensus"
repeat_region       23017..23847
                    /note="L1M4 repeat: matches 3686..4511 of consensus"
repeat_region       23830..24512
                    /note="L1MB8 repeat: matches 4523..5213 of consensus"
repeat_region       24786..24823
                    /note="19 copies 2 mer ta 79% conserved"
repeat_region       24824..24897
                    /note="37 copies 2 mer at 77% conserved"
repeat_region       24934..25258
                    /note="L1M47 repeat: matches 5867..6210 of consensus"
repeat_region       25801..26039
                    /note="L1M4 repeat: matches 4177..4427 of consensus"
repeat_region       26661..26941
                    /note="L2 repeat: matches 1972..2246 of consensus"
repeat_region       26942..27233
                    /note="AluJo repeat: matches 1..290 of consensus"
repeat_region       27234..27556
                    /note="L2 repeat: matches 2246..2588 of consensus"
repeat_region       27557..28145
                    /note="L1R1 repeat: matches 203..785 of consensus"
repeat_region       28140..28392
                    /note="L1R1 repeat: matches 1..252 of consensus"
repeat_region       28395..28643
                    /note="HUERS-P2 repeat: matches 2833..3086 of consensus"
repeat_region       28720..29694
                    /note="HUERS-P2 repeat: matches 1850..2831 of consensus"
repeat_region       29695..30235
                    /note="L1M4 repeat: matches 740..1323 of consensus"
repeat_region       30235..30553
```





## RESULT 4

AL136445  
 LOCUS Human DNA sequence from clone RP3-376F14 on chromosome 6q12-13, complete sequence.  
 DEFINITION  
 ACCESSION AL136445  
 VERSION AL136445.9 GI:9369006  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 51877)  
 Direct Submission  
 Submitted (20-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk  
 On Jul 22, 2000 this sequence version replaced gi:9367907.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr6  
 RP3-376F14 is from the library RPCI-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/  
 VECTOR: pCIPAC2  
 IMPORTANT: This sequence is not the entire insert of clone RP3-376F14. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RP3-376F14 is at 1 in this sequence. The true left end of clone RPI-257B21 is at 51778 in this sequence. The true right end of clone RPI-138F4 is at 185 in this sequence.

FEATURES  
 source  
 Location/Qualifiers  
 1..51877  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="q12-13"  
 /clone="RP3-376F14"  
 /clone\_lib="RPCI-3"  
 complement(1..158)  
 /note="match: STS: Em:HS138F4S"  
 321..633  
 /note="AluSp repeat: matches 1..311 of consensus"  
 760..791  
 /note="8 copies 4 mer agag 90% conserved"  
 1251..1450  
 /note="L1ME1 repeat: matches 5687..5897 of consensus"  
 1471..1514  
 /note="22 copies 2 mer at 75% conserved"  
 1810..1873  
 /note="4 copies 16 mer 93% conserved"  
 2056..2267

repeat\_region  
 2944..3095  
 /note="MER20 repeat: matches 1..217 of consensus"  
 repeat\_region  
 4172..4745  
 /note="MIR repeat: matches 34..198 of consensus"  
 repeat\_region  
 4742..5769  
 /note="L2 repeat: matches 523..1136 of consensus"  
 repeat\_region  
 5819..5842  
 /note="L2 repeat: matches 1337..2505 of consensus"  
 repeat\_region  
 5882..5945  
 /note="12 copies 2 mer tg 100% conserved"  
 repeat\_region  
 5894..5953  
 /note="MIR repeat: matches 188..258 of consensus"  
 repeat\_region  
 6050..6134  
 /note="L2 repeat: matches 2690..2750 of consensus"  
 repeat\_region  
 6354..6443  
 /note="MIR repeat: matches 168..256 of consensus"  
 repeat\_region  
 7315..7625  
 /note="L2 repeat: matches 2622..2706 of consensus"  
 repeat\_region  
 8848..8978  
 /note="AluX repeat: matches 1..308 of consensus"  
 repeat\_region  
 9104..9395  
 /note="t1MB4 repeat: matches 6050..6180 of consensus"  
 repeat\_region  
 10714..10755  
 /note="AluJb repeat: matches 1..303 of consensus"  
 misc\_feature  
 complement(11564..12033)  
 /note="21 copies 2 mer aa 76% conserved"  
 repeat\_region  
 12480..12919  
 /note="match: GSS: Em:AQ730254"  
 repeat\_region  
 12920..13214  
 /note="L2 repeat: matches 170..595 of consensus"  
 repeat\_region  
 13215..14653  
 /note="AluX repeat: matches 1..303 of consensus"  
 repeat\_region  
 14987..15458  
 /note="L2 repeat: matches 595..2108 of consensus"  
 repeat\_region  
 16236..16452  
 /note="L2 repeat: matches 2244..2749 of consensus"  
 repeat\_region  
 16458..16918  
 /note="MER30 repeat: matches 2..230 of consensus"  
 misc\_feature  
 /note="match: GSS: Em:B94425"  
 misc\_feature  
 16464..16907  
 /note="match: GSS: Em:AQ086190"  
 repeat\_region  
 20376..20446  
 /note="match: GSS: Em:AQ086190"  
 misc\_feature  
 22227..22921  
 /note="L2 repeat: matches 2635..2709 of consensus"  
 misc\_feature  
 22872..23171  
 /note="match: GSS: Em:AQ894591"  
 misc\_feature  
 /note="match: GSS: Em:AQ814633"  
 repeat\_region  
 23468..23751  
 /note="MUT1A1 repeat: matches 73..352 of consensus"  
 misc\_feature  
 24297..24578  
 /note="match: GSS: Em:B89604"  
 repeat\_region  
 24856..24972  
 /note="match: GSS: Em:B89604"  
 misc\_feature  
 /note="MER53 repeat: matches 1..189 of consensus"  
 complement(25074..25570)  
 /note="match: GSS: Em:AQ761253"  
 repeat\_region  
 25785..26028  
 /note="match: GSS: Em:AQ761253"  
 repeat\_region  
 26745..27053  
 /note="LIPB3 repeat: matches 5896..6136 of consensus"  
 repeat\_region  
 27054..27405  
 /note="MER46C repeat: matches 5..323 of consensus"  
 repeat\_region  
 28058..28113  
 /note="THE1C repeat: matches 1..369 of consensus"  
 repeat\_region  
 /note="28 copies 2 mer tg 80% conserved"  
 misc\_feature  
 complement(29766..30453)  
 /note="match: GSS: Em:AQ475294"  
 misc\_feature  
 complement(30079..30444)  
 /note="match: GSS: Em:AQ596956"  
 misc\_feature  
 complement(30087..30283)  
 /note="match: GSS: Em:AQ415484"  
 misc\_feature  
 30447..30861  
 /note="match: GSS: Em:AQ371434"  
 repeat\_region  
 30980..31005  
 /note="match: GSS: Em:AQ371434"  
 repeat\_region  
 /note="13 copies 2 mer tc 92% conserved"

```
misc_feature 31798..32313
/note="match: GSS: Em:AQ457417"
repeat_region 32424..32979
/note="MLT2B repeat: matches 1..448 of consensus"
repeat_region 33011..33525
/note="L1PA7 repeat: matches 5624..6137 of consensus"
repeat_region 33813..33864
/note="2 copies 26 mer 94% conserved"
misc_feature complement(34633..35115)
/note="match: GSS: Em:AQ685318"
repeat_region 34778..34952
/note="MIR repeat: matches 71..251 of consensus"
repeat_region 35034..35081
/note="12 copies 4 mer at 75% conserved"
repeat_region 35241..35268
/note="14 copies 2 mer 92% conserved"
misc_feature 36239..36638
/note="match: GSS: Em:B62948"
repeat_region 38552..38855
/note="AluX repeat: matches 1..304 of consensus"
misc_feature complement(38995..39275)
/note="match: GSS: Em:AQ076337"
repeat_region 39033..39102
/note="MER21B repeat: matches 723..792 of consensus"
repeat_region 39124..39882
/note="MER21B repeat: matches 1..794 of consensus"
misc_feature complement(39815..40597)
/note="match: GSS: Em:AQ894592"
misc_feature 40637..41167
/note="match: GSS: Em:AQ665820"
misc_feature 41363..41826
/note="match: GSS: Em:AQ711725"
repeat_region 41498..41781
/note="AluB repeat: matches 19..302 of consensus"
repeat_region 41820..42375
/note="L1PA16 repeat: matches 5593..6152 of consensus"
repeat_region 42689..42732
/note="11 copies 4 mer ctca 88% conserved"
repeat_region 43083..43139
/note="MER91B repeat: matches 36..89 of consensus"
misc_feature complement(44702..45088)
/note="match: GSS: Em:B89306"
misc_feature match: STS: Em:G61198"
misc_feature 46934..47217
/note="match: GSS: Em:AQ164004"
misc_feature 46960..47426
/note="match: GSS: Em:AQ205505"
repeat_region 47052..47371
/note="MER2 repeat: matches 2..334 of consensus"
repeat_region 47574..47700
/note="MIR repeat: matches 20..145 of consensus"
repeat_region 48219..48329
/note="MER5A repeat: matches 4..114 of consensus"
repeat_region 49415..49622
/note="MIR repeat: matches 29..241 of consensus"
repeat_region 50974..51162
/note="L1MD2 repeat: matches 5695..5891 of consensus"
repeat_region 51425..51877
Query Match 2.08; Score 21; DB 11; Length 51877;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 541 agatacagaattattcttggg 561
|||||
Db 19491 AGATACAGAAATTATTCTTGGG 19511

RESULT 5
HS657J8/c HS657J8 98954 bp DNA PRI 23-NOV-1999
LOCUS Human DNA sequence from clone 657J8 on chromosome Xq26.1-26.3
DEFINITION Contains GSS, complete sequence.
```

```

repeat_region 11059..11277
/note="t1M4 repeat: matches 5489..5711 of consensus"
repeat_region 11507..11580
/note="37 copies 2 mer aa 65% conserved"
repeat_region 11589..11740
/note="t1ME repeat: matches 529..670 of consensus"
repeat_region 11741..11938
/note="MER63A repeat: matches 5..210 of consensus"
repeat_region 14939..14623
/note="t1ME repeat: matches 670..3159 of consensus"
repeat_region 14624..14897
/note="AluSx repeat: matches 5..297 of consensus"
repeat_region 14898..15235
/note="t1ME repeat: matches 3159..3475 of consensus"
repeat_region 15246..17091
/note="t1ME repeat: matches 3640..5663 of consensus"
repeat_region 17090..17370
/note="t1ME repeat: matches 5880..6161 of consensus"
repeat_region 18592..24735
/note="t1PA2 repeat: matches 7..6144 of consensus"
repeat_region 25257..25450
/note="MIR repeat: matches 22..212 of consensus"
repeat_region 25722..25863
/note="L2 repeat: matches 2563..2748 of consensus"
repeat_region 25879..26256
/note="L2 repeat: matches 2391..2749 of consensus"
repeat_region 26369..26455
/note="L2 repeat: matches 2637..2745 of consensus"
repeat_region 26656..26827
/note="L1P4 repeat: matches 5627..5809 of consensus"
repeat_region 26842..27224
/note="MLT1J repeat: matches 1..458 of consensus"
repeat_region 27239..27385
/note="MER33 repeat: matches 2..153 of consensus"
repeat_region 27598..27665
/note="34 copies 2 mer aa 68% conserved"
misc_feature complement(<29625..30042)
/note="match: GSS AQ024839"
repeat_region 29950..30169
/note="L1PA13 repeat: matches 5937..6156 of consensus"
repeat_region 30171..33705
/note="L1P repeat: matches 1888..5333 of consensus"
repeat_region 34229..34329
/note="L2 repeat: matches 2655..2748 of consensus"
repeat_region 34553..35112
/note="MLT2B repeat: matches 1..553 of consensus"
repeat_region 35708..35819
/note="L1P repeat: matches 2473..2597 of consensus"
repeat_region 35853..36134
/note="L1PA13 repeat: matches 651..370 of consensus"
repeat_region 36135..36451
/note="MER61A repeat: matches 4..354 of consensus"
repeat_region 36452..36750
/note="L1PA13 repeat: matches 370..71 of consensus"
repeat_region 36919..37025
/note="L1PA16 repeat: matches 652..751 of consensus"
repeat_region 37266..40505
/note="L1PA16 repeat: matches 752..3902 of consensus"
repeat_region 40506..40867
/note="THE1B repeat: matches 3..364 of consensus"
repeat_region 40868..41364
/note="THE1B-INTERNAL repeat: matches 1..497 of consensus"
repeat_region 41365..41688
/note="AluY repeat: matches 1..303 of consensus"
repeat_region 41669..42752
/note="THE1B-INTERNAL repeat: matches 497..1580 of consensus"
repeat_region 42753..43118
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 43119..45286
/note="L1PA16 repeat: matches 3902..6157 of consensus"
repeat_region 46599..46861
/note="MLT1B repeat: matches 233..405 of consensus"

```

```

repeat_region 46862..47156
/note="AluSq repeat: matches 1..306 of consensus"
repeat_region 47157..47404
/note="MLT1B repeat: matches 4..233 of consensus"
repeat_region 47499..47540
/note="21 copies 2 mer ac 98% conserved"
repeat_region 47922..47988
/note="MLT2A repeat: matches 1..84 of consensus"
repeat_region 47989..50715
/note="HERVL repeat: matches 391..3161 of consensus"
repeat_region 50716..51159
/note="MLT2B repeat: matches 1..444 of consensus"
repeat_region 51160..53667
/note="HERVL repeat: matches 3161..5654 of consensus"
repeat_region 53668..54176
/note="MLT2A repeat: matches 1..449 of consensus"
misc_feature complement(54195..54782)
/note="match: GSS B59069 clone 2014M20"
misc_feature complement(54335..54643)
/note="match: GSS AQ226070"
repeat_region 56557..56643
/note="L1PA3 repeat: matches 6060..6146 of consensus"
repeat_region 56992..57039
/note="24 copies 2 mer ac 98% conserved"
repeat_region 57260..57587
/note="L1MA8 repeat: matches 5942..6285 of consensus"
repeat_region 58280..58830
/note="TIGER1 repeat: matches 1881..2418 of consensus"
repeat_region 58846..59620
/note="L1PA15 repeat: matches 5383..6157 of consensus"
repeat_region 59794..59952
/note="L2 repeat: matches 2586..2750 of consensus"
repeat_region 61761..61930
/note="L2 repeat: matches 2578..2750 of consensus"
repeat_region 63300..63327
/note="14 copies 2 mer ga 100% conserved"
repeat_region 64328..64981
/note="L1PA16 repeat: matches 5513..6157 of consensus"
repeat_region 64982..65286
/note="AluSg repeat: matches 1..304 of consensus"
repeat_region 65287..66691
/note="L1PA16 repeat: matches 4085..5513 of consensus"
repeat_region 66673..67634
/note="MST-INTERNAL repeat: matches 536..1520 of consensus"
repeat_region 67635..68180
/note="MLT2D repeat: matches 6..553 of consensus"
repeat_region 68189..68606
/note="MST-INTERNAL repeat: matches 126..541 of consensus"
repeat_region 68622..68836
/note="MST repeat: matches 1..226 of consensus"

```

```

Query Match 2.0%; Score 21; DB 46; Length 98954;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 634 tcaagtaaatgcttaattat 654
|||||
DB 46157 TCAGTAAATGCTTAATTAT 46137

```

```

RESULT 6
AC006945 AC006945 120309 bp DNA ROD 15-JUN-2000
LOCUS Mus musculus chromosome 6 clone ct7-b453113 strain 129/Sv ES cell
DEFINITION line C37, complete sequence.
ACCESSION AC006945
VERSION AC006945.14 GI:8567816
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

```

REFERENCE
AUTHORS      1 (bases 1 to 120309)
TITLE        Chen, F., Do, T., Do, A., McDermid, H. and Roe, B.A.
JOURNAL      Mus musculus Chromosome 6 BAC Clone b453113 In CES Region
REFERENCE
AUTHORS      2 (bases 1 to 120309)
TITLE        Unpublished
JOURNAL
AUTHORS      Chen, F., Do, T., Do, A., McDermid, H. and Roe, B.A.
TITLE        Direct Submission
JOURNAL      Submitted (02-MAR-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      3 (bases 1 to 120309)
TITLE        Chen, F., Do, T., Do, A., McDermid, H. and Roe, B.A.
JOURNAL      Direct Submission
              Submitted (07-JUL-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      4 (bases 1 to 120309)
TITLE        Chen, F., Do, T., Do, A., McDermid, H. and Roe, B.A.
JOURNAL      Direct Submission
              Submitted (24-JUL-1999) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      5 (bases 1 to 120309)
TITLE        Chen, F., Do, T., Do, A., McDermid, H. and Roe, B.A.
JOURNAL      Direct Submission
              Submitted (15-FEB-2000) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      6 (bases 1 to 120309)
TITLE        Chen, F., Do, T., Do, A., McDermid, H. and Roe, B.A.
JOURNAL      Direct Submission
              Submitted (22-FEB-2000) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      7 (bases 1 to 120309)
TITLE        Chen, F., Do, T., Do, A., McDermid, H. and Roe, B.A.
JOURNAL      Direct Submission
              Submitted (15-JUN-2000) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
COMMENT      On Jun 15, 2000 this sequence version replaced gi:7019301.
FEATURES
source
1..120309
/organism="Mus musculus"
/strain="129/Sv EV cell line CJ7"
/db_xref="taxon:10090"
/chromosome="6"
/clone="ct7-b453113"
/clone_lib="CitBCJ7 mouse BAC library"
BASE COUNT 30828 a 29967 c 30215 g 29299 t
ORIGIN

Query Match      2.08; Score 21; DB 88; Length 120309;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 690 acatgcagctgtctctgtga 710
|||||
Db 89098 ACAATGCAGCTGTCTCTGTGA 89118

RESULT 7
AC004093/c AC004093 157684 bp DNA ROD 31-JAN-1998
DEFINITION Mus musculus (129SV) DNA, unmapped BAC 10817, complete sequence.
ACCESSION AC004093
VERSION AC004093.1 GI:2828220
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS      1 (bases 1 to 157684)
TITLE        Lamerdin, J.E., McCready, P.M., Skowronski, E., Adamson, A.W.,
              Burkhardt-Schultz, K., Gordon, L., Kyle, A., Ramirez, M., Stilwagen, S.,
              Barnes, J., Danganan, L., Poundstone, P., Christensen, M.,
              Georgescu, A., Avila, J., Liu, S., Bruce, R., Quan, G., Montgomery, M.,
              Ow, D., Nollan, M., Trong, S., Kobayashi, A., Olsen, A.O. and
              Carrano, A.V.
TITLE        Untitled
JOURNAL      Unpublished
REFERENCE
AUTHORS      2 (bases 1 to 157684)
TITLE        Lamerdin, J.E.
JOURNAL      Direct Submission
              Submitted (31-JAN-1998) Joint Genome Institute, Lawrence Livermore
              National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
FEATURES
source
1..157684
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="BAC-10817"
/cell_line="CJ7"
/lab_host="HS996 (modified DH10B/r)"
/note="BAC clone obtained from Genome Systems. Library
made from cell line CJ7, derived from mouse strain 129SV."
complement(1..646)
/rpt_family="L1"
repeat_region 649..802
repeat_region 676..753
misc_feature 676..753
/note="Predicted exon, program: graill2exons_human_1.3,
frame: 0, quality: excellent, score: 86.000"
repeat_region 1516..1563
/rpt_family="(GGA)n"
repeat_region complement(3180..3208)
/rpt_family="POLY_A"
misc_feature complement(4735..4991)
/note="Predicted exon, program: graill2exons_human_1.3,
frame: 1, quality: good, score: 52.000"
misc_feature complement(5770..6199)
/note="BLASTN similarity to U17089 (72..503); match: 0.93,
score: 8.5e-139; database searched: nt; Mus musculus MT
transposon-like element, clone Mt17."
repeat_region complement(6121..6200)
/rpt_family="THE1-INTERNAL"
repeat_region 9771..9823
/rpt_family="(CA)n"
misc_feature complement(9823..9908)
/note="Predicted exon, program: graill2exons_human_1.3,
frame: 2, quality: excellent, score: 86.000"
repeat_region 10150..10173
/rpt_family="POLY_A"
repeat_region complement(10838..10877)
/rpt_family="AT-rich"
misc_feature complement(11272..11610)
/note="Predicted exon, program: graill2exons_human_1.3,
frame: 1, quality: good, score: 51.000"
repeat_region complement(12688..12820)
/rpt_family="ELAM_C"
repeat_region complement(13313..13358)
/rpt_family="AT-rich"
repeat_region complement(13865..13969)
/rpt_family="L1MC4"
repeat_region complement(14440..14884)
/rpt_family="L1MD3"
misc_feature 14681..14844
/note="Predicted exon, program: graill2exons_human_1.3,
frame: 1, quality: good, score: 67.000"
repeat_region complement(15549..15632)
/rpt_family="AluJo"
repeat_region complement(16026..16058)
/rpt_family="(CAAA)n"
repeat_region complement(16233..16292)
/rpt_family="(CA)n"
repeat_region 16292..16419

```

```

/rpt_family="GA)n"
complement(117364..17432)
/rpt_family="AT_rich"
18848..18901
/rpt_family="(CA)n"
18906..19029
/rpt_family="(TAAA)n"
19108..19156
/rpt_family="AluJb"
complement(19158..19272)
/rpt_family="(TAAA)n"
19627..19758
/rpt_family="(TAGA)n"
complement(20957..21319)
/Note="Ribosomal pseudogene.-DPS similarity to
gi1132881|sp|P12947|RL31_HUMAN 60S RIBOSOMAL PROTEIN L31;
gi1171328|pir|IR5HU31 ribosomal protein L31 - human;
gi1171329|pir|IR5RT31 ribosomal protein L31 - rat; gi136130
(X15940) ribosomal protein L31 (AA 1-125) [Homo
sapiens]; gi157115 (X04809) ribosomal protein L31 (AA
1-125) [Rattus norvegicus]; gi11632855|gnl|P1D1e2|76436
(281183) ribosomal protein L31 [Sus scrofa]. Score: 405
Identity: 91/125 (73%)."
22316..22474
/rpt_family="AluJb"
23291..23335
/rpt_family="Alu"
23316..23350
/Note="Unresolved number of A's in homopolymeric A-
region."
23336..23434
/rpt_family="POLY_A"
23728..23867
/rpt_family="AluJo"
23874..23925
/rpt_family="(GA)n"
24688..24798
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 52.000"
24883..25000
/rpt_family="(GA)n"
complement(25229..25300)
/rpt_family="7SLRNA"
complement(25355..25449)
/rpt_family="7SLRNA"
complement(25390..25465)
/rpt_family="AluSp"
25477..25637
/rpt_family="AluSg/x"
27395..27440
/rpt_family="(CA)n"
27440..27585
/Note="Low confidence in sequence within low complexity
repeat region."
27441..27566
/rpt_family="(CA)n"
27598..27742
/rpt_family="(CA)n"
complement(28223..28355)
/rpt_family="AluSp"
complement(30491..30518)
/rpt_family="AT_rich"
complement(30815..30844)
/rpt_family="AT_rich"
31335..31401
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 53.000"
31599..31720
/rpt_family="(GA)n"
32391..32519
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 56.000"
complement(33730..33788)
/rpt_family="L1MB8"
complement(33815..33874)
/rpt_family="L1MB8"
complement(34206..34568)
/rpt_family="L1M4"
34803..34858
/rpt_family="AluSc"
34827..34923
/rpt_family="FLAM_A"
complement(34928..35353)
/rpt_family="L1"
35362..35535
/rpt_family="L1M4"
35462..35731
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: good, score: 74.000"
complement(36417..36506)
/rpt_family="POLY_A"
36874..36894
/rpt_family="AT_rich"
complement(37247..37346)
/rpt_family="(GA)n"
complement(37348..37505)
/rpt_family="FRAM"
38772..39106
/Note="BLASTN similarity to Y08837 (1..335); match: 0.8,
score: 1.1e-82; database searched: nt; H.sapiens mRNA for
RAD51-like protein"
38773..39400
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 65.000"
complement(41066..41217)
/rpt_family="FRAM"
41459..41479
/rpt_family="AT_rich"
41769..41823
/rpt_family="AT_rich"
complement(43283..43334)
/rpt_family="AluJ/monomer"
complement(43492..43523)
/rpt_family="(CAAA)n"
44327..44447
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 72.000"
44390..44438
Query Match 2.0%; Score 21; DB 88; Length 157684;
Best Local Similarity 100.0%; Pred.No.1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 693 tgcagcttgcctgtgtgtg 713
|||||
Db 95197 TGCAGCTTGCTGTGTGTG 95177

RESULT 8
AC067861/c AC067861 180978 bp DNA HTG 09-JUL-2000
LOCUS Homo sapiens chromosome 8 clone RP11-351M15 map 8, WORKING DRAFT
DEFINITION SEQUENCE, 25 unordered pieces.
ACCESSION AC067861
VERSION AC067861.4 GI:8980931
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 180978)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-351M15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 180978)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

```

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,  
 Campopiano, A., Castle, A., Choce, P., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Domano, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Hesford, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lechoczy, J.,  
 Levine, R., Liew, C., Liu, G., Locke, K., Macdonald, P., Marguis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Melrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Testaye, S., Theodores, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
 Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 9, 2000 this sequence version replaced gi:8576272.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

## TITLE

## JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L9211

Center clone name: 351.M.15

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Assembly: Dye-terminator Big Dye; 100% of reads

Consensus quality: 163664 bases at least Q40

Consensus quality: 171960 bases at least Q30

Consensus quality: 175688 bases at least Q20

Insert size: 182000; agarose-gel

Insert size: 178578; sum-of-contigs

Quality coverage: 3.8 in Q20 bases; agarose-gel

Quality coverage: 3.8 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 25 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1386: contig of 1386 bp in length  
 1387 1486: gap of 100 bp  
 1487 3312: contig of 1826 bp in length  
 3313 3412: gap of 100 bp  
 3413 6028: contig of 2616 bp in length  
 6029 6128: gap of 100 bp  
 6129 9208: contig of 3080 bp in length  
 9209 9308: gap of 100 bp  
 9309 11748: contig of 2440 bp in length  
 11749 11848: gap of 100 bp  
 11849 12786: contig of 938 bp in length  
 12787 12886: gap of 100 bp  
 12887 15072: contig of 2186 bp in length  
 15073 15172: gap of 100 bp  
 15173 18295: contig of 3123 bp in length  
 18296 18395: gap of 100 bp  
 18396 21952: contig of 3557 bp in length  
 21953 22052: gap of 100 bp

\* 22053 25389: contig of 3337 bp in length  
 25390 25489: gap of 100 bp  
 25490 28734: contig of 3245 bp in length  
 28735 28834: gap of 100 bp  
 28835 34041: contig of 5207 bp in length  
 34042 34141: gap of 100 bp  
 34142 38263: contig of 4122 bp in length  
 38264 38363: gap of 100 bp  
 38364 44707: contig of 6344 bp in length  
 44708 44807: gap of 100 bp  
 44808 51727: contig of 6920 bp in length  
 51728 51827: gap of 100 bp  
 51828 59668: contig of 7841 bp in length  
 59669 59768: gap of 100 bp  
 59769 64938: contig of 5170 bp in length  
 64939 65038: gap of 100 bp  
 65039 72833: contig of 7795 bp in length  
 72834 72933: gap of 100 bp  
 72934 81393: contig of 8460 bp in length  
 81394 81493: gap of 100 bp  
 81494 90782: contig of 9289 bp in length  
 90783 90882: gap of 100 bp  
 90883 101839: contig of 10957 bp in length  
 101840 101939: gap of 100 bp  
 101940 115221: contig of 13282 bp in length  
 115222 115321: gap of 100 bp  
 115322 129777: contig of 14456 bp in length  
 129778 129877: gap of 100 bp  
 129878 154359: contig of 24482 bp in length  
 154360 154459: gap of 100 bp  
 154460 180978: contig of 26519 bp in length.

## FEATURES

## Source

Location/Qualifiers  
 1..180978  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone.lib="RPC1-11 Human Male BAC"  
 /clone.lib="RPC1-11 Human Male BAC"  
 1..1386  
 /note="assembly\_fragment"  
 1487..3312  
 /note="assembly\_fragment"  
 3413..6028  
 /note="assembly\_fragment"  
 6129..9208  
 /note="assembly\_fragment"  
 9309..11748  
 /note="assembly\_fragment"  
 11849..12786  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 12887..15072  
 /note="assembly\_fragment"  
 15173..18295  
 /note="assembly\_fragment"  
 18396..21952  
 /note="assembly\_fragment"  
 22053..25389  
 /note="assembly\_fragment"  
 25490..28734  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:right  
 28835..34041  
 /note="assembly\_fragment"  
 34142..38263  
 /note="assembly\_fragment"  
 38364..44707  
 /note="assembly\_fragment"  
 44808..51727  
 /note="assembly\_fragment"

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

COMMENT

On Aug 11, 2000 this sequence version replaced gl:9438582.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HAZG  
Center clone name: RP11-141C22  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-terminator Big Dye; 88% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 162956 bases at least Q40  
Consensus quality: 178252 bases at least Q30  
Consensus quality: 185100 bases at least Q20  
Estimated insert size: 176369; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 3.4x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 39 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

1 49943: contig of 49943 bp in length  
49544 50043: gap of unknown length  
50044 50703: contig of 10660 bp in length  
50704 60803: gap of unknown length  
60804 75186: contig of 14383 bp in length  
75187 75286: gap of unknown length  
75287 81935: contig of 6649 bp in length  
81936 82035: gap of unknown length  
82036 89332: contig of 7197 bp in length  
89333 96305: gap of unknown length  
96306 96405: gap of unknown length  
96406 102567: contig of 6162 bp in length  
102568 102667: gap of unknown length  
102668 108430: contig of 5763 bp in length  
108431 108530: gap of unknown length  
108531 114680: contig of 6150 bp in length  
114681 114780: gap of unknown length  
114781 119819: contig of 5039 bp in length  
119820 119919: gap of unknown length  
119920 125700: contig of 5781 bp in length  
125701 125800: gap of unknown length  
125801 130582: contig of 4762 bp in length  
130583 130662: gap of unknown length  
130663 135648: contig of 4986 bp in length  
135649 135748: gap of unknown length  
135749 140221: contig of 4473 bp in length  
140222 140321: gap of unknown length  
140322 143112: contig of 2791 bp in length  
143113 143212: gap of unknown length  
143213 147531: contig of 4319 bp in length  
147532 147631: gap of unknown length  
147632 151489: contig of 3858 bp in length  
151490 151589: gap of unknown length  
151590 154960: contig of 3371 bp in length  
154961 155060: gap of unknown length  
155061 157209: contig of 2149 bp in length  
157210 157309: gap of unknown length  
157310 159422: contig of 2113 bp in length  
159423 159522: gap of unknown length  
159523 162233: contig of 2711 bp in length  
162234 162333: gap of unknown length

51828 .59668  
/note="assembly\_fragment"  
59769 .64938  
/note="assembly\_fragment"  
65039 .72833  
/note="assembly\_fragment"  
72934 .81393  
/note="assembly\_fragment"  
81494 .90782  
/note="assembly\_fragment"  
90883 .101839  
/note="assembly\_fragment"  
101940 .115221  
/note="assembly\_fragment"  
115322 .129777  
/note="assembly\_fragment"  
129878 .154359  
/note="assembly\_fragment"  
154460 .180978  
/note="assembly\_fragment"  
53732 a 34332 c 33667 g 56836 t 2411 others  
BASE COUNT  
ORIGIN

Query Match 2.0%; Score 21; DB 55; Length 180978;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 541 agatcacagattattcttggga 561  
|||||  
DB 86389 AGATACAGATTATCTTGGGA 86369

RESULT 9  
AC067720/c  
LOCUS  
DEFINITION Homo sapiens chromosome 3 clone RP11-141C22, WORKING DRAFT  
SEQUENCE, 39 unordered pieces.  
AC067720  
AC067720.9 GI:9795440  
VERSION  
KEYWORDS HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Muzny, D.M., Adams, C., Bailey, M., Barberia, J., Blankenburg, K.,  
Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,  
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,  
David, R., Delgado, O., Deshazo, D., Ding, Y., Donah-Rashid, N.,  
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,  
Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,  
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hogue, M.,  
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,  
Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,  
Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,  
Morris, S., Nash, S., Nelson, A., Nguyen, C., Nguyen, N., Nguyen, S.,  
Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,  
Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,  
Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R.,  
Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M.,  
Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,  
Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and  
Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 193998)  
Worley, K.C.  
Direct Submission  
Submitted (27-APR-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

```
* 162334 164482: contig of 2149 bp in length
* 164483 164582: gap of unknown length
* 164583 167867: contig of 3285 bp in length
* 167868 167967: gap of unknown length
* 167968 170137: contig of 2170 bp in length
* 170138 170237: gap of unknown length
* 170238 172037: contig of 1800 bp in length
* 172038 172137: gap of unknown length
* 172138 174008: contig of 1871 bp in length
* 174009 174108: gap of unknown length
* 174109 175941: contig of 1833 bp in length
* 175942 176041: gap of unknown length
* 176042 178267: contig of 2226 bp in length
* 178268 178367: gap of unknown length
* 178368 180031: contig of 1664 bp in length
* 180032 180131: gap of unknown length
* 180132 181352: contig of 1221 bp in length
* 181353 181452: gap of unknown length
* 181453 183004: contig of 1552 bp in length
* 183005 183105: gap of unknown length
* 183106 184197: contig of 1093 bp in length
* 184198 184297: gap of unknown length
* 184298 185697: contig of 1400 bp in length
* 185698 185797: gap of unknown length
* 185798 185894: contig of 1197 bp in length
* 185895 187094: gap of unknown length
* 187095 188318: contig of 1224 bp in length
* 188319 188418: gap of unknown length
* 188419 189814: contig of 1396 bp in length
* 189815 189914: gap of unknown length
* 189915 191273: contig of 1359 bp in length
* 191274 191373: gap of unknown length
* 191374 192786: contig of 1413 bp in length
* 192787 192886: gap of unknown length
* 192887 193998: contig of 1112 bp in length.
FEATURES
    source
        1..19398
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /clone="RP11-141C22"
        1..49943
            /note="assembly_name:Contig54"
        50044..60703
            /note="assembly_name:Contig53"
            clone_end:T7
            vector_side:right
        60804..75186
            /note="assembly_name:Contig52"
        75287..81935
            /note="assembly_name:Contig51"
        82036..89232
            /note="assembly_name:Contig50"
        89333..96305
            /note="assembly_name:Contig49"
        96406..102567
            /note="assembly_name:Contig48"
        102668..108430
            /note="assembly_name:Contig47"
        108531..114680
            /note="assembly_name:Contig46"
        114781..119819
            /note="assembly_name:Contig45"
        119920..125700
            /note="assembly_name:Contig44"
        125801..130562
            /note="assembly_name:Contig43"
        130663..135648
            /note="assembly_name:Contig42"
        135749..140221
            /note="assembly_name:Contig41"
        140322..143112
            /note="assembly_name:Contig40"
```

```
misc_feature 143213..147531
    /note="assembly_name:Contig39"
misc_feature 147632..151489
    /note="assembly_name:Contig38"
misc_feature 151590..154960
    /note="assembly_name:Contig37"
misc_feature 155061..157209
    /note="assembly_name:Contig36"
    clone_end:SP6
    vector_side:left
misc_feature 157310..159422
    /note="assembly_name:Contig35"
misc_feature 159523..162233
    /note="assembly_name:Contig34"
misc_feature 162334..164482
    /note="assembly_name:Contig33"

Query Match      2.0%; Score 21; DB 55; Length 193998;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 gatggagtagaactgtctga 782
    |||||
Db 95085 GATGGAGTAGAACTGTCTGA 95065

RESULT 10
XLU95097
LOCUS XLU95097 1447 bp mRNA VRT 05-MAY-1997
DEFINITION Xenopus laevis mitotic phosphoprotein 43 mRNA, partial cds.
ACCESSION U95097
VERSION U95097.1 GI:2072293
KEYWORDS African clawed frog.
SOURCE Xenopus laevis
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
    Xenopodidae; Xenopus.
REFERENCE 1 (bases 1 to 1447)
AUTHORS Stukenberg,P.T., Lustig,K.D., McGarry,T.J., King,R.W., Kuang,J. and
    Kirschner,M.W.
TITLE Systematic identification of mitotic phosphoproteins
JOURNAL Curr. Biol. 7 (5), 338-348 (1997)
MEDLINE 9727270
REFERENCE 2 (bases 1 to 1447)
AUTHORS Stukenberg,P.T., Lustig,K.D., McGarry,T.J., King,R.W., Jian,K. and
    Kirschner,M.W.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-1997) Cell Biology, Harvard Medical School, 25
    Shattuck St, Boston, MA 02115, USA
COMMENT This is the partial sequence of a cDNA isolated in a screen for
    mitotic phosphoproteins
FEATURES
    source
        1..1447
            /organism="Xenopus laevis"
            /db_xref="taxon:8355"
            <1..1170
            /note="MP43"
            /codon_start=1
            /product="mitotic phosphoprotein 43"
            /protein_id="AAC60121.1"
            /db_xref="GI:2072294"
            /translation="PIDSNSARAEDVKGENRFTYQGVQKHKLQPSHGNTTMDL
                ENNFNSNRLRMELDAIKKYEQHSNYQVISTLEGDSLQKAVRDQKQYRELEQ
                RDLFSENRLRMELDAIKKYEQHSNYQVISTLEGDSLQKAVRDQKQYRELEQ
                ANDDLERAKRATIMLESDEORLQAIERNAFLENDELKENLLESVQRKLRELEQ
                CELAVQKQKPKSNMVSPTERTDTSVQASIAISAPITPLSQRCASLTPLSPFSR
                TSLDDGYSGLPLTPCARIKALNIVGDLIRKVGALSKLASCRNFVHEQSPNRLTSSV
                ARMKTRGLENLRSIASGSSVSGVGLIKRLEFGSPSNTPMQGMHSPQGVKMI"
BASE COUNT 454 a 292 c 341 g 360 t
ORIGIN
```



Query Match 1.9%; Score 20; DB 4; Length 1447;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 973 cagaacataaacattact 992  
 DB 76 CAGAACATACATTTACT 95

RESULT 11  
 AC005173/c 38754 bp DNA PRI 24-JUN-1998  
 LOCUS Homo sapiens clone UWGC:gl564a009 from 7p14-15, complete sequence.  
 DEFINITION AC005173  
 ACCESSION AC005173  
 VERSION AC005173.1 GI:3252831  
 KEYWORDS HTG.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 38754)  
 Iadonato, S.P., Yu, J., Wong, G.K.-S., Magness, C.L., Green, E.D.,  
 Green, P. and Olson, M.V.  
 Large-scale MCD Mapping and Sequencing of Human Chromosome 7  
 Unpublished (1998)  
 2 (bases 1 to 38754)  
 Iadonato, S.P., Yu, J., Wong, G.K.-S., Magness, C.L., Green, E.D.,  
 Green, P. and Olson, M.V.  
 Direct Submission  
 Submitted (24-JUN-1998) Human Genome Center, University of  
 Washington, Box 352145, Seattle, WA 98195, USA  
 University of Washington Human Genome Center  
 Box 352145 Seattle, WA 98195  
 Contact: Shawn Iadonato (iadonato@u.washington.edu)  
 5': UWGC:gl564a016  
 3': UWGC:gl564a072

Overlapping Sequences:  
 -----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 Genbank flat file format but are available as part  
 of this entry's ASN.1 file.  
 -----  
 Double stranded (DS) coverage: 100.0%  
 DS or two chemistry coverage: 100.0%  
 Single stranded regions: 0

Sequence Validation:  
 This sequence has been validated by Multiple Complete Digest  
 Mapping. Comparison of the experimentally derived map digest  
 fragments with sequence-predicted fragments is given below.  
 Small fragments below a variable cutoff (approximately 400-600bp)  
 are not mapped and hence do not appear in the table. There are no  
 significant remaining discrepancies between the experimental and  
 predicted values. Uniquely ordered fragment groups are separated  
 by dashed lines.

Map	Seq	Map	Seq	Map	Seq
959.33	955.00	716.50	718.00	4292.57	4282.00
2554.17	2568.00	1986.11	1995.00	3795.00	3790.00
1296.43	1297.00	1722.00	1709.00	2616.25	2641.00
705.67	689.00	1274.83	1275.00	1098.50	1117.00
1209.50	1220.00	1479.17	1471.00	4649.75	4705.00

FEATURES	Location/Qualifiers
source	1. .38754
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7p14-15"
	/clone="HGRI:YWSS1564"
	/sub_clone="UWGC:gl564a009"
	/cell_line="GM10791"
	/clone_lib="E. Green Chromosome 7 YAC Resource"
repeat_region	552..999
	/rpt_family="MER4"
repeat_region	1120..1631
	/rpt_family="MER1"
repeat_region	2313..2377
	/rpt_family="MIR"
repeat_region	2604..2822
	/rpt_family="MER7"
STS	3536..3631
	/standard_name="HUMSW1489"
	/note="Genbank Accession: G00382"
repeat_region	complement(3641..3780)
	/rpt_family="MIR"
repeat_region	complement(4598..4725)
	/rpt_family="MIR"
repeat_region	4830..4969
	/rpt_family="MIR"
repeat_region	7200..7780
	/rpt_family="MLT1"
repeat_region	7378..7591
	/rpt_family="Alu"
repeat_region	complement(10516..10747)
	/rpt_family="MER33"
repeat_region	11406..11676
	/rpt_family="Alu"
repeat_region	14306..15270
	/rpt_family="MER42"
repeat_region	complement(16266..16518)
	/rpt_family="MER7"
repeat_region	16844..16774
	/rpt_family="MER42"
repeat_region	19199..19582
	/rpt_family="MER25"
repeat_region	22078..22627
	/rpt_family="L1"
repeat_region	22692..22951





## MAPPING INFORMATION:

This clone was provided to the Washington University Genome Sequencing Center by Dr. Stephen Scherer, Department of Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada. This clone was isolated as part of a chromosome 7 mapping effort supported by the Canadian Genome Analysis and Technology program. For more information, see <http://www.genet.sickkids.on.ca/chromosome7>

## SOURCE INFORMATION:

This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).

Cell line: lymphoblastoid

Haplotypes: two

VECTOR: pBelOABAC

Selection: chloramphenicol

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is GS1-440B14, 200 bp overlap. The actual start of this clone is at base position 17859 of GS1-440B14.

This clone contains SNS's SWS3975 (NID:gl408140) and SWS2937 (NID:gl113649).

## FEATURES

source	Location/Qualifiers
	1. 118595
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/clone="GS1-16514"
	/map_lib="GSBAC1"
	/map="7q21"
repeat_region	complement(1133..1283)
repeat_region	/rpt_family="MER"
repeat_region	complement(1451..1712)
repeat_region	/rpt_family="MER"
repeat_region	1534..2684
repeat_region	/rpt_family="MER"
repeat_region	complement(2463..2753)
repeat_region	/rpt_family="MER"
repeat_region	5618..5896
repeat_region	/rpt_family="ALU"
repeat_region	complement(7920..7954)
repeat_region	/rpt_family="L1"
repeat_region	9219..9509
repeat_region	/rpt_family="ALU"
repeat_region	complement(9618..9679)
repeat_region	/rpt_family="L1"
repeat_region	complement(10507..10531)
repeat_region	/rpt_family="L1"
repeat_region	complement(10532..10831)
repeat_region	/rpt_family="ALU"
repeat_region	15209..15320
repeat_region	/rpt_family="ALU"
repeat_region	complement(15617..15754)
repeat_region	/rpt_family="L1"
repeat_region	complement(16510..16902)
repeat_region	/rpt_family="L1"
repeat_region	complement(17144..17336)
repeat_region	/rpt_family="L1"
repeat_region	complement(17384..17608)
repeat_region	/rpt_family="L1"
repeat_region	complement(17869..17973)
repeat_region	/rpt_family="L1"
repeat_region	complement(19141..19158)
repeat_region	/rpt_family="L1"
repeat_region	complement(21271..21293)
repeat_region	/rpt_family="L1"
repeat_region	complement(23037..23320)
repeat_region	/rpt_family="ALU"
misc_feature	24472..24539
repeat_region	/note="match to human EST T75033 (NID:g691795) yc86a02.r1"
	27485..27952
	/rpt_family="L1"
repeat_region	28229..28293
	/rpt_family="L1"
repeat_region	28311..28602
	/rpt_family="ALU"
repeat_region	complement(29073..29549)
	/rpt_family="L1"
misc_feature	30754..30847
	/note="match to human EST T75033 (NID:g691795) yc86a02.r1"
misc_feature	33124..33305
	/note="match to human EST T75033 (NID:g691795) yc86a02.r1"
misc_feature	33188..33305
	/note="match to human EST D61451 (NID:g970206)"
misc_feature	36119..36219
	/note="match to human EST D61451 (NID:g970206)"
misc_feature	36154..36238
	/note="match to human EST T78741 (NID:g697250) yd01e05.r1"
repeat_region	38471..38488
	/rpt_family="L1"
misc_feature	38489..38569
	/note="match to human EST T78741 (NID:g697250) yd01e05.r1"
repeat_region	40728..40751
	/rpt_family="L1"
repeat_region	complement(42383..42424)
	/rpt_family="L1"
repeat_region	44564..44758
	/rpt_family="ALU"
repeat_region	44787..44861
	/rpt_family="ALU"
repeat_region	46198..46234
	/rpt_family="L1"
repeat_region	46263..46292
	/rpt_family="L1"
repeat_region	complement(47659..47959)
	/rpt_family="ALU"
repeat_region	49614..49905
	/rpt_family="ALU"
misc_feature	50906..51028
	/note="match to human EST T78741 (NID:g697250) yd01e05.r1"
repeat_region	complement(51404..51929)
	/rpt_family="L1"
repeat_region	51898..53777
	/rpt_family="L1"
repeat_region	56330..56373
	/rpt_family="L1"
repeat_region	56652..56733
	/rpt_family="L1"
repeat_region	57076..57259
	/rpt_family="L1"
repeat_region	57643..57692
	/rpt_family="L1"
repeat_region	57768..58067
	/rpt_family="ALU"
repeat_region	58083..58277
	/rpt_family="L1"
repeat_region	58723..59057
	/rpt_family="L1"
repeat_region	59102..59385
	/rpt_family="L1"
repeat_region	59961..60237
	/rpt_family="L1"
misc_feature	61408..61499
	/note="match to human EST N36698 (NID:gl157840) yx92e10.r1"
misc_feature	61463..61499
	/note="similar to human EST R17730 (NID:g771340) yg04e02.r1"
misc_feature	61463..61499
	/note="similar to human EST R35924 (NID:g792825) yg68c05.r1"
repeat_region	complement(62873..62912)
	/rpt_family="L1"
repeat_region	complement(63820..64107)
	/rpt_family="ALU"

us-09-101-423a-1.olg.rge

```

repeat_region 66613..66653
/rpt_family="L1"
misc_feature 68723..68872
/note="match to human EST N36698 (NID:gil157840)"
misc_feature 68723..68872

Query Match 1.9%; Score 20; DB 9; Length 118595;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 gccctccctctccagtc 42
|||||
Db 76409 GCCTCTCCCTCTCCAGTC 76390

RESULT 14
AC026395/c
LOCUS AC026395 152044 bp DNA HTG 15-JUN-2000
DEFINITION Homo sapiens chromosome 10 clone RP11-45D20, WORKING DRAFT
SEQUENCE, 44 unordered pieces.
AC026395
ACCESSION AC026395.3 GI:8567738
VERSION 24099
KEYWORDS HTG; HTGS_PRASEL; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152044)
Smith,D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequencing Data
Unpublished
2 (bases 1 to 152044)
Smith,D.R.
Direct Submission
Submitted (22-MAR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On Jun 15, 2000 this sequence version replaced gi:7330305.
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg344
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembled program: Phrap; version 990315
Consensus quality: 126801 bases at least Q40
Consensus quality: 138138 bases at least Q30
Consensus quality: 140422 bases at least Q20
Insert size: 147744; sum-of-contigs
Quality coverage: 3.2x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1006: contig of 1006 bp in length
* 1007 1106: gap of unknown length
* 1107 2475: Contig of 1369 bp in length
* 2476 2575: gap of unknown length
* 2576 3927: contig of 1352 bp in length
* 3928 4027: gap of unknown length
* 4028 5333: contig of 1306 bp in length
* 5334 5433: gap of unknown length
* 5434 5564: contig of 1131 bp in length
* 5564: gap of unknown length
* 5565 6664: gap of unknown length
* 6665 7870: contig of 1206 bp in length
* 7871 9384: contig of 1414 bp in length
* 9385 9485: gap of unknown length
* 9485 10791: contig of 1307 bp in length
* 10791 10891: gap of unknown length
* 10891 13010: contig of 2119 bp in length
* 13010 13111: gap of unknown length
* 13111 14617: contig of 1507 bp in length
* 14618 14717: gap of unknown length
* 14718 16217: contig of 1500 bp in length
* 16218 16318: gap of unknown length
* 16318 18107: contig of 1789 bp in length
* 18107 18206: gap of unknown length
* 18207 19498: contig of 1292 bp in length
* 19499 19598: gap of unknown length
* 19599 20867: contig of 1269 bp in length
* 20868 20967: gap of unknown length
* 20968 22203: contig of 1236 bp in length
* 22204 24098: contig of 1795 bp in length
* 24099 24198: gap of unknown length
* 24199 26005: contig of 1807 bp in length
* 26006 26105: gap of unknown length
* 26106 27441: contig of 1336 bp in length
* 27442 27541: gap of unknown length
* 27542 29037: contig of 1496 bp in length
* 29038 29137: gap of unknown length
* 29138 30660: contig of 1523 bp in length
* 30661 30760: gap of unknown length
* 30761 32755: contig of 1995 bp in length
* 32756 32855: gap of unknown length
* 32856 34483: contig of 1628 bp in length
* 34484 34583: gap of unknown length
* 34584 37578: contig of 2995 bp in length
* 37579 37678: gap of unknown length
* 37679 39699: contig of 2021 bp in length
* 39700 39799: gap of unknown length
* 39800 42411: contig of 2612 bp in length
* 42412 42511: gap of unknown length
* 42512 45638: contig of 3127 bp in length
* 45639 45738: gap of unknown length
* 45739 49739: contig of 4001 bp in length
* 49740 49839: gap of unknown length
* 49840 54276: contig of 4437 bp in length
* 54277 54376: gap of unknown length
* 54377 59051: contig of 4675 bp in length
* 59052 59151: gap of unknown length
* 59152 64626: contig of 5475 bp in length
* 64627 64726: gap of unknown length
* 64727 67988: contig of 3262 bp in length
* 67989 68088: gap of unknown length
* 68089 72495: contig of 4406 bp in length
* 72496 72595: gap of unknown length
* 72596 75643: contig of 3049 bp in length
* 75644 75743: gap of unknown length
* 75744 79510: contig of 3767 bp in length
* 79511 79610: gap of unknown length
* 79611 84664: contig of 5054 bp in length
* 84665 84764: gap of unknown length
* 84765 90514: contig of 5750 bp in length
* 90515 90614: gap of unknown length
* 90615 96185: contig of 5571 bp in length
* 96186 96285: gap of unknown length
* 96286 102048: contig of 5763 bp in length
* 102049 102148: gap of unknown length
* 102149 108361: contig of 6213 bp in length
* 108362 108461: gap of unknown length
* 108462 114329: contig of 5868 bp in length
* 114330 114429: gap of unknown length
* 114430 120644: contig of 6215 bp in length
* 120645 120744: gap of unknown length

```

\* 120745 128025: contig of 7281 bp in length  
 \* 128026 128125: gap of unknown length  
 \* 128126 138788: contig of 10663 bp in length  
 \* 138789 138888: gap of unknown length  
 \* 138889 152044: contig of 13156 bp in length.

FEATURES  
 source  
 1. 152044  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-45D20"  
 /clone\_lib="RPCI-11"

BASE COUNT 40176 a 33928 c 33229 g 40362 t 4349 others  
 ORIGIN

Query Match 1.9% Score 20; DB 51; Length 152044;  
 Best Local Similarity 100.0%; Pred. No. 5.9;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 892 cagcttttcactgacaaaca 911  
 |||||  
 Db 80334 CAGCTTTTCACTGACAAACA 80315

## RESULT 15

AC025947 157057 bp DNA HTG 10-JUN-2000  
 LOCUS Homo sapiens chromosome Chromosome 10 clone RP11-78A18, WORKING  
 DEFINITION DRAFT SEQUENCE, 26 unordered pieces.

AC025947  
 AC025947.3 GI:8439851  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.

## SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 157057)

## REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

Submitted (18-MAR-2000) Genome Therapeutics Corporation, 100 Beaver  
 Street, Waltham, MA 02453, USA  
 On Jun 10, 2000 this sequence version replaced gi:7528340.

## COMMENT

----- genome center  
 Center: Genome Therapeutics Corporation  
 Center code: GTC  
 Web site: <http://www.genomecorp.com/>  
 Contact: [gtc-seqcenter@genomecorp.com](mailto:gtc-seqcenter@genomecorp.com)  
 ----- Project Information  
 ----- Summary Statistics  
 Sequencing vector: N/A  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 990315  
 Consensus quality: 140586 bases at least Q40  
 Consensus quality: 148596 bases at least Q30  
 Consensus quality: 150061 bases at least Q20  
 Insert size: 154557; sum-of-contigs  
 Quality coverage: 3.7x in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 26 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1544: contig of 1544 bp in length  
 \* 1644: gap of unknown length  
 \* 1645: contig of 1574 bp in length  
 \* 3218: gap of unknown length  
 \* 3219 5022: contig of 1704 bp in length  
 \* 5023 5122: gap of unknown length  
 \* 5123 6553: contig of 1431 bp in length  
 \* 6554 8651: gap of unknown length  
 \* 8652 8751: gap of unknown length  
 \* 8752 10843: contig of 2092 bp in length  
 \* 10844 10943: gap of unknown length  
 \* 10944 13772: contig of 2829 bp in length  
 \* 13773 13872: gap of unknown length  
 \* 13873 15825: contig of 1953 bp in length  
 \* 15826 15925: gap of unknown length  
 \* 15926 15991: contig of 3666 bp in length  
 \* 15992 19691: gap of unknown length  
 \* 19692 22350: contig of 2659 bp in length  
 \* 22351 22450: gap of unknown length  
 \* 22451 25948: contig of 3498 bp in length  
 \* 25949 28048: gap of unknown length  
 \* 28049 29641: contig of 3593 bp in length  
 \* 29642 29741: gap of unknown length  
 \* 29742 34106: contig of 4364 bp in length  
 \* 34107 34205: gap of unknown length  
 \* 34206 38484: contig of 4279 bp in length  
 \* 38485 38584: gap of unknown length  
 \* 38585 41992: contig of 3408 bp in length  
 \* 41993 42092: gap of unknown length  
 \* 42093 47920: contig of 5828 bp in length  
 \* 47921 48020: gap of unknown length  
 \* 48021 54416: contig of 6396 bp in length  
 \* 54417 54516: gap of unknown length  
 \* 54517 60194: contig of 5678 bp in length  
 \* 60195 60294: gap of unknown length  
 \* 60295 67639: contig of 7345 bp in length  
 \* 67640 67739: gap of unknown length  
 \* 67740 74790: contig of 7051 bp in length  
 \* 74791 74890: gap of unknown length  
 \* 74891 85631: contig of 10741 bp in length  
 \* 85632 85731: gap of unknown length  
 \* 85732 93940: contig of 8209 bp in length  
 \* 93941 94040: gap of unknown length  
 \* 94041 104191: contig of 10151 bp in length  
 \* 104192 104291: gap of unknown length  
 \* 104292 116388: contig of 12096 bp in length  
 \* 116389 116487: gap of unknown length  
 \* 116488 135857: contig of 19370 bp in length  
 \* 135858 135957: gap of unknown length  
 \* 135958 157057: contig of 21100 bp in length.

## FEATURES

## source

1. 157057  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="Chromosome 10"  
 /clone="RP11-78A18"  
 /clone\_lib="RPCI-11"

BASE COUNT 45178 a 32550 c 33372 g 44448 t 2509 others  
 ORIGIN

Query Match 1.9% Score 20; DB 50; Length 157057;  
 Best Local Similarity 100.0%; Pred. No. 5.9;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 892 cagcttttcactgacaaaca 911  
 |||||

Db 152909 CAGCTTTTCACTGACAAACA 152928

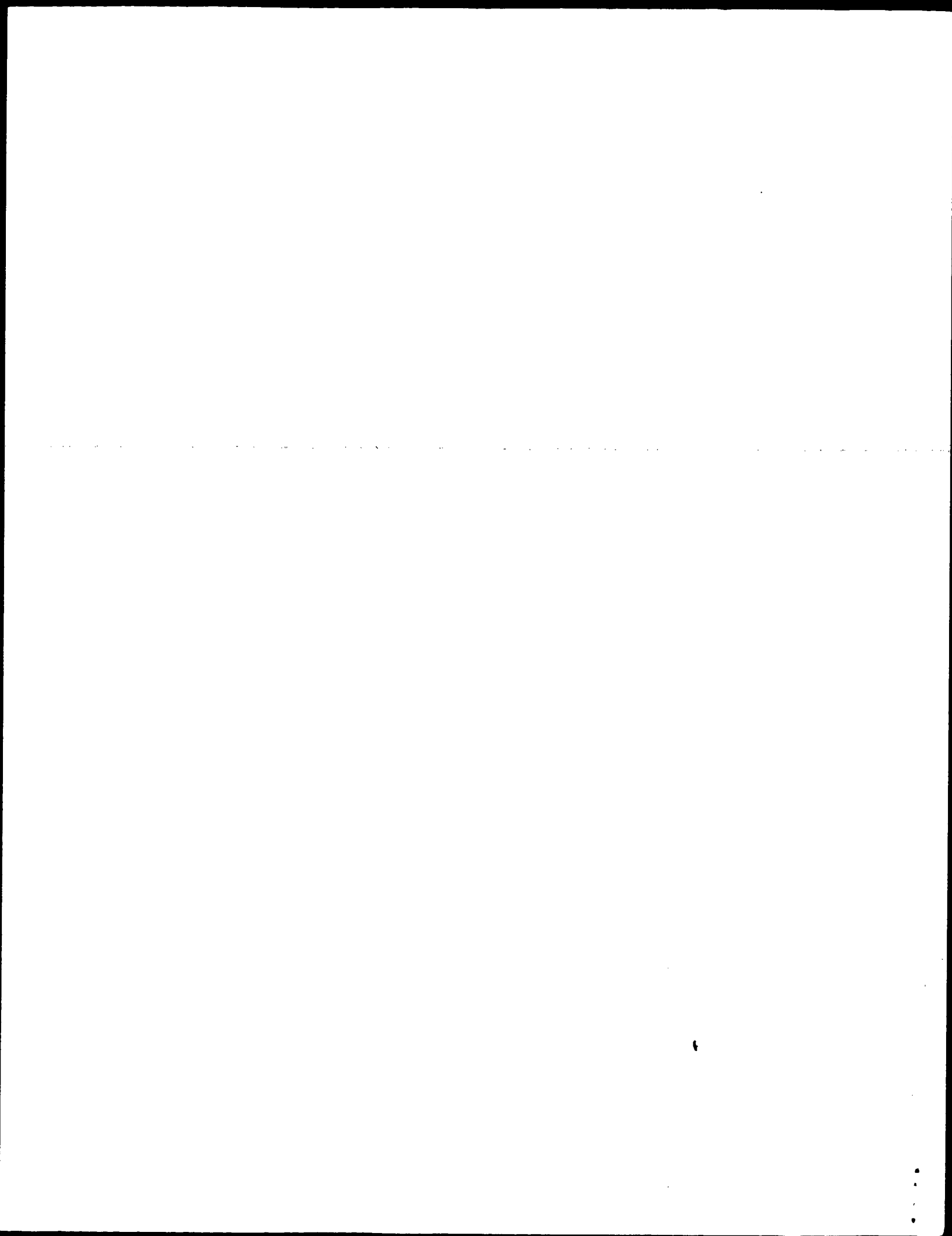
Search completed: November 4, 2000, 03:36:43

Mon Nov 6 10:14:00 2000

us-09-101-423a-1.olig.rge

Page 19

Job time: 19713 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: November 4, 2000, 00:32:16 ; Search time 255.06 Seconds  
(without alignments)  
1521.444 Million cell updates/sec

Title: US-09-101-423A-1  
Perfect score: 1033  
Sequence: 1 ctctctgggtctatgtc.....gtagacataactccatggt 1033  
Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 480022 seqs, 187831343 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 960044  
Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Listing first 45 summaries

Database :		N_Geneseq_36:*	
1:	/SID56/gcgdata/geneseq/geneseq/NA1980.DAT:*	13	17
2:	/SID56/gcgdata/geneseq/geneseq/NA1981.DAT:*	14	17
3:	/SID56/gcgdata/geneseq/geneseq/NA1982.DAT:*	15	17
4:	/SID56/gcgdata/geneseq/geneseq/NA1983.DAT:*	16	17
5:	/SID56/gcgdata/geneseq/geneseq/NA1984.DAT:*	17	17
6:	/SID56/gcgdata/geneseq/geneseq/NA1985.DAT:*	18	16
7:	/SID56/gcgdata/geneseq/geneseq/NA1986.DAT:*	19	16
8:	/SID56/gcgdata/geneseq/geneseq/NA1987.DAT:*	20	16
9:	/SID56/gcgdata/geneseq/geneseq/NA1988.DAT:*	21	16
10:	/SID56/gcgdata/geneseq/geneseq/NA1989.DAT:*	22	16
11:	/SID56/gcgdata/geneseq/geneseq/NA1990.DAT:*	23	16
12:	/SID56/gcgdata/geneseq/geneseq/NA1991.DAT:*	24	16
13:	/SID56/gcgdata/geneseq/geneseq/NA1992.DAT:*	25	16
14:	/SID56/gcgdata/geneseq/geneseq/NA1993.DAT:*	26	16
15:	/SID56/gcgdata/geneseq/geneseq/NA1994.DAT:*	27	16
16:	/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:*	28	16
17:	/SID56/gcgdata/geneseq/geneseq/NA1996.DAT:*	29	16
18:	/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:*	30	16
19:	/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:*	31	16
20:	/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:*	32	16
21:	/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:*	33	16

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES		Query		Description	
Result No.	Score	Match	Length	ID	
1	862	83.4	1034	18	T72785
c	2	17	1.6	231	20
c	3	17	1.6	304	21
c	4	17	1.6	514	21
c	5	17	1.6	1020	17
c	6	17	1.6	1079	17
c	7	17	1.6	1100	19
c	8	17	1.6	1340	20
c	9	17	1.6	1521	20
c	10	17	1.6	1716	17
c	11	17	1.6	1830	13
c	12	17	1.6	1900	10

13	17	1.6	2940	18	T90501
c	14	1.6	3203	21	Z49226
c	15	1.6	3691	18	T90500
c	16	1.6	4147	18	T89334
c	17	1.6	4870	18	T60301
c	18	1.6	20	19	V31186
c	19	1.5	28	19	X09790
c	20	1.5	49	21	Z59078
c	21	1.5	80	20	Z09575
c	22	1.5	150	19	X10511
c	23	1.5	240	21	Z61722
c	24	1.5	287	16	T22929
c	25	1.5	306	16	T21132
c	26	1.5	327	20	V86788
c	27	1.5	344	20	V87708
c	28	1.5	349	15	Q70066
c	29	1.5	378	21	Z80107
c	30	1.5	405	20	Z11177
c	31	1.5	420	21	Z42664
c	32	1.5	435	20	V86374
c	33	1.5	437	15	Q70063
c	34	1.5	510	21	Z40677
c	35	1.5	543	20	X20858
c	36	1.5	701	20	Z11184
c	37	1.5	933	20	V81471
c	38	1.5	1000	16	T01268
c	39	1.5	1098	20	V84594
c	40	1.5	1132	19	V56045
c	41	1.5	1183	20	Z27252
c	42	1.5	1245	14	Q41102
c	43	1.5	1245	19	V35607
c	44	1.5	1254	16	T00621
c	45	1.5	1322	8	N70148

ALIGNMENTS

RESULT 1	
T72785	
ID T72785 standard; DNA; 1034 BP.	
XX	
AC T72785;	
XX	
DT 22-SEP-1997 (first entry)	
DE Metastasis inducing DNA C2.	
XX	
KW Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;	
KW osteopontin; ss.	
XX	
OS Homo sapiens.	
XX	
PN WO9725443-A1.	
XX	
PD 17-JUL-1997.	
XX	
PF 10-JAN-1997; 97WO-GB000074.	
XX	
PR 10-JAN-1996; 96GB-0000470.	
XX	
PA (UWLI-) UNIV LIVERPOOL.	
XX	
PI Barracough BR, Rudland PS;	
XX	
DR WPI; 1997-372878/34.	
XX	
PT New isolated metastasis-inducing DNA - used to develop products to	
PT identify and treat patients at risk from metastatic tumours	
XX	
PS Claim 8; Page 25; 38pp; English.	
XX	
CC Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12	



XX SQ Sequence 231 BP; 76 A; 64 C; 45 G; 46 T; 0 other;

Query Match 1.6%; Score 17; DB 20; Length 231;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 gacagactctgggacag 80  
|||||  
DB 180 GACAGACTCTGGGACAG 164

RESULT 3  
Z49261/C  
ID Z49261 standard; cDNA; 304 BP.

XX AC Z49261;

XX DT 07-MAR-2000 (first entry)

XX DE Human hydrolase homologue HHH-4 cDNA fragment 3773551H1.

XX KW Hydrolase; homologue; HHH-4; N-terminal asparagine amidohydrolase;  
XX KW vanin-1; glycosyl hydrolase; glucosyl hydrolase;  
XX KW N-acetylglucosamine 6-P deacetylase; diagnosis;  
XX KW treatment; prevention; expression; disorder; carbohydrate; metabolism;  
XX KW antagonist; reproductive disorder; cell proliferation; ds.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO9961626-A2.

XX PD 02-DEC-1999.

XX PF 28-MAY-1999; 99WO-US12021.

XX PR 29-MAY-1998; 98US-0087236.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Bandman O, Hillman JL, Yue H, Lal P, Corley NC, Guegler KJ;  
XX PI Patterson C, Baughn MR;

XX DR WPI; 2000-062716/05.

XX PT New human hydrolase homologues, useful for treating or preventing a  
XX PT carbohydrate metabolism disorder -

XX PS Disclosure; Page 87; 91pp; English.

XX CC Sequences Z49232, Z49234, Z49236, Z49238, Z49241, Z49243, Z49244,  
XX CC Z49245, Z49246, Z49253, Z49257, and Z49261 represent human hydrolase  
XX CC homologue HHH-4 cDNA fragments. The full-length cDNA sequence is given  
XX CC in Z49226. Nucleic acids encoding HHH-4 were initially identified in a  
XX CC neonatal dermal microvascular endothelial cell cDNA library. HHH-4 has  
XX CC homology to glycosyl hydrolase (BL00775). Glycosyl hydrolases cleave the  
XX CC glycosidic bond between two carbohydrates or between a  
XX CC carbohydrate and a non-carbohydrate moiety, and are frequently  
XX CC associated with disorders of carbohydrate metabolism. The invention  
XX CC relates to human hydrolase homologues HHH-1 to HHH-7 (Y58165-Y58171)  
XX CC which respectively have homology to N-terminal asparagine amidohydrolase,  
XX CC vanin-1, glycosyl hydrolases, glucosyl hydrolase and N-acetylglucosamine 6-P  
XX CC deacetylase. Such homologues are useful in methods for diagnosing,  
XX CC treating or preventing disorders associated with expression of  
XX CC hydrolases. The hydrolase homologues are useful for treating or  
XX CC preventing a carbohydrate metabolism disorder. Antagonists of these  
XX CC hydrolases can be used to treat or prevent a reproductive or cell  
XX CC proliferation disorder.

XX SQ Sequence 304 BP; 77 A; 75 C; 100 G; 52 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 304;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 tgcctctccctctctcc 38  
|||||  
DB 178 TGCCTCTCCCTCTCTCC 162

RESULT 4  
Z49244/C  
ID Z49244 standard; cDNA; 514 BP.

XX AC Z49244;

XX DT 07-MAR-2000 (first entry)

XX DE Human hydrolase homologue HHH-4 cDNA fragment 1657538F6.

XX KW Hydrolase; homologue; HHH-4; N-terminal asparagine amidohydrolase;  
XX KW vanin-1; glycosyl hydrolase; glucosyl hydrolase;  
XX KW N-acetylglucosamine 6-P deacetylase; diagnosis;  
XX KW treatment; prevention; expression; disorder; carbohydrate; metabolism;  
XX KW antagonist; reproductive disorder; cell proliferation; ds.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO9961626-A2.

XX PD 02-DEC-1999.

XX PF 28-MAY-1999; 99WO-US12021.

XX PR 29-MAY-1998; 98US-0087236.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Bandman O, Hillman JL, Yue H, Lal P, Corley NC, Guegler KJ;  
XX PI Patterson C, Baughn MR;

XX DR WPI; 2000-062716/05.

XX PT New human hydrolase homologues, useful for treating or preventing a  
XX PT carbohydrate metabolism disorder -

XX PS Disclosure; Page 80; 91pp; English.

XX CC Sequences Z49232, Z49234, Z49236, Z49238, Z49241, Z49243, Z49244,  
XX CC Z49245, Z49246, Z49253, Z49257, and Z49261 represent human hydrolase  
XX CC homologue HHH-4 cDNA fragments. The full-length cDNA sequence is given  
XX CC in Z49226. Nucleic acids encoding HHH-4 were initially identified in a  
XX CC neonatal dermal microvascular endothelial cell cDNA library. HHH-4 has  
XX CC homology to glycosyl hydrolase (BL00775). Glycosyl hydrolases cleave the  
XX CC glycosidic bond between two carbohydrates or between a  
XX CC carbohydrate and a non-carbohydrate moiety, and are frequently  
XX CC associated with disorders of carbohydrate metabolism. The invention  
XX CC relates to human hydrolase homologues HHH-1 to HHH-7 (Y58165-Y58171)  
XX CC which respectively have homology to N-terminal asparagine amidohydrolase,  
XX CC vanin-1, glycosyl hydrolases, glucosyl hydrolase and N-acetylglucosamine 6-P  
XX CC deacetylase. Such homologues are useful in methods for diagnosing,  
XX CC treating or preventing disorders associated with expression of  
XX CC hydrolases. The hydrolase homologues are useful for treating or  
XX CC preventing a carbohydrate metabolism disorder. Antagonists of these  
XX CC hydrolases can be used to treat or prevent a reproductive or cell  
XX CC proliferation disorder.

XX SQ Sequence 514 BP; 112 A; 121 C; 177 G; 93 T; 11 other;

Query Match 1.6%; Score 17; DB 21; Length 514;  
Best Local Similarity 100.0%; Pred. No. 38;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 tgcctctcccttctcc 38  
|||||

Db 111 TGCCTCTCCCTTCTCC 95

## RESULT 5

ID T30293/C  
T30293 standard; cDNA; 1020 BP.

XX AC T30293;

XX DT 15-OCT-1996 (first entry)

XX DE Human SLAM3 cDNA clone pSECSlam.

XX KW SLAM3; SLAM3; pSECSlam; cell surface antigen; T lymphocyte;  
XX KW T-cell co-stimulatory molecule; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 61..978

FT sig\_peptide /\*tag= a

FT /\*tag= b

FT mat\_peptide 142..975

FT /\*tag= c

FT variation 761

FT /\*tag= d

FT /\*note= "splice junction where the transmembrane  
domain sequence of SLAM1 is deleted is  
at position 761"

XX PN WO9617060-A1.

XX PD 06-JUN-1996.

XX PF 29-NOV-1995; 95WO-US14942.

XX PR 07-JUN-1995; 95US-0481777.

XX PR 02-DEC-1994; 94US-0348792.

XX PA (SCHE ) SCHERING CORP.

XX PI Aversa G, Chang C, Cocks BG, De Vries JB;

XX DR WPI; 1996-277780/28.

XX DR P-PSDB; R97630.

XX PT New isolated SLAM protein and gene - used to develop prods. for use  
XX PT in diagnostics, screening studies and therapy involving  
XX PT immunological responses

XX PS Claim 10; Page 72-73; 94pp; English.

XX CC A cDNA clone (T30293), designated pSECSlam, codes for SLAM3  
XX CC (R97630), a secreted form of human SLAM that lacks a transmembrane  
XX CC region. SLAM3 is a novel T-cell stimulatory molecule which,  
XX CC when engaged, potentiates T-cell expansion and induces a Th10/Th1  
XX CC cytokine prodn. profile. pSECSlam was isolated from a T-cell cDNA  
XX CC library by expression cloning using anti-SLAM monoclonal antibody  
XX CC A12 for selection. Other human SLAM cDNA clones (see also  
XX CC T30291-92 and T30294) were also isolated. The SLAM gene was  
XX CC mapped to 1q21.3 and 1q22, a region correlating to that for the  
XX CC gene for systemic lupus erythematosus susceptibility. The cDNA  
XX CC can be used for prodn. of recombinant SLAM1, or as a probe or  
XX CC primer.

XX SQ Sequence 1020 BP; 285 A; 283 C; 242 G; 210 T; 0 other;

Query Match 1.6%; Score 17; DB 17; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 38;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 gacagactctggacag 80  
|||||

Db 924 GACAGACTCTGGACAG 908

## RESULT 6

ID T30294/C  
T30294 standard; cDNA; 1079 BP.

XX AC T30294;

XX DT 15-OCT-1996 (first entry)

XX DE Human SLAM4 cDNA clone pCYTslam.

XX KW SLAM4; SLAM4; pCYTslam; cell surface antigen; T lymphocyte;  
XX KW T-cell co-stimulatory molecule; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT variation 1..145

FT /\*tag= a

FT /\*note= "an alternative exon at the 5' end (up  
to nucleotide 145) predicts that SLAM4  
lacks a leader sequence"

FT CDS 153..1076

FT /\*tag= b

FT mat\_peptide 153..1073

FT /\*tag= c

XX PN WO9617060-A1.

XX PD 06-JUN-1996.

XX PF 29-NOV-1995; 95WO-US14942.

XX PR 07-JUN-1995; 95US-0481777.

XX PR 02-DEC-1994; 94US-0348792.

XX PA (SCHE ) SCHERING CORP.

XX PI Aversa G, Chang C, Cocks BG, De Vries JE;

XX DR WPI; 1996-277780/28.

XX DR P-PSDB; R97631.

XX PT New isolated SLAM protein and gene - used to develop prods. for use  
XX PT in diagnostics, screening studies and therapy involving  
XX PT immunological responses

XX PS Claim 10; Page 75-77; 94pp; English.

XX CC A cDNA clone (T30294), designated pCYTslam, codes for human  
XX CC SLAM4 (R97631), a cytoplasmic form of a novel T-cell stimulatory  
XX CC molecule which, when engaged, potentiates T-cell expansion and  
XX CC induces a Th10/Th1 cytokine prodn. profile. It was isolated from a  
XX CC T-cell cDNA library by expression cloning using anti-SLAM monoclonal  
XX CC antibody A12 for selection. Other human SLAM cDNA clones (see also  
XX CC T30291-93) were also isolated. The SLAM gene was mapped to a  
XX CC region close to 1q21.3 and 1q22; this region correlates to the gene  
XX CC for systemic lupus erythematosus susceptibility. The cDNA can be  
XX CC used for prodn. of recombinant SLAM4, or as a probe or primer.

XX SQ Sequence 1079 BP; 311 A; 271 C; 262 G; 235 T; 0 other;

Query Match 1.6%; Score 17; DB 17; Length 1079;  
Best Local Similarity 100.0%; Pred. No. 38;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 gacagactctgggacag 80  
|||||  
Db 1022 GACAGACTCTGGGACAG 1006

RESULT 7  
V18890/c  
ID V18890 standard; cDNA; 1100 BP.  
XX  
AC V18890;  
XX  
DT 09-JUN-1998 (first entry)  
XX  
DE Rat Hypertension related calcium regulated gene (HcARG) cDNA.  
XX  
KW Hypertension related calcium regulated gene; HcARG; rat parathyroid;  
KW extracellular calcium concentration; antibody; hypertension; ss;  
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;  
KW cancer; inflammatory disease; asthma.  
XX  
OS Rattus rattus.  
XX

Key Location/Qualifiers  
FH 131..806  
CDS /\*tag= a  
FT /product= "Hypertension related calcium regulator"  
FT 173..196  
FT misc\_feature /\*tag= b  
FT /note= "EF-hand like motif"

XX WO9749807-A2.  
XX 31-DEC-1997.  
XX 23-JUN-1997; 97WO-CA00439.  
XX 21-JUN-1996; 96US-0567495.  
XX (GOSS/) GOSSARD F.  
XX (HAME/) HAMEY P.  
XX (LEWA/) LEWANCZUK R.  
XX (TREM/) TREMBLAY J.  
XX Gossard F, Hamet P, Lewanczuk R, Tremblay J;  
XX WPI; 1998-077171/07.  
XX P-PSDB; W37723.  
XX Hypertension related calcium regulated gene - useful to develop  
XX products to treat or detect, e.g. hypertension, stroke,  
XX osteoporosis, heart failure, cancer, diabetes or asthma  
XX  
XX Claim 2; Pages 25-26; 46pp; English.  
XX This nucleic acid sequence encodes the hypertension related calcium  
XX regulated gene (HcARG), which was isolated from the rat parathyroid.  
XX Its expression is regulated by extracellular calcium concentration.  
XX An antibody against the protein, can be used to detect or modulate  
XX (e.g. enhance or inhibit) abnormal calcium levels. They can  
XX specifically be used to detect or treat, e.g. hypertension,  
XX hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,  
XX inflammatory disease, and asthma.  
XX  
XX Sequence 1100 BP; 260 A; 279 C; 315 G; 246 T; 0 other;

Query Match 1.6%; Score 17; DB 19; Length 1100;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 929 ctgagattataaactgt 945

Db 987 CTGAGATTAAACTGT 971  
|||||

RESULT 8  
Z42034/c  
ID Z42034 standard; cDNA; 1340 BP.  
XX  
AC Z42034;  
XX  
DT 31-JAN-2000 (first entry)  
XX  
DE Human endometrium tumour cDNA derived EST 54.

XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;  
KW treatment; uterine; gene therapy; expressed sequence tag; ss.  
XX  
OS Homo sapiens.  
XX  
PN DE19817948-A1.  
XX  
PD 21-OCT-1999.

XX 17-APR-1998; 98DE-1017948.  
XX 17-APR-1998; 98DE-1017948.  
XX (META-) METAGEN-GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;  
XX WPI; 1999-591957/51.  
XX P-PSDB; Y60101, Y60102, Y60103.  
XX New nucleic acid sequences expressed in uterine cancer tissues, and  
XX derived polypeptides, for treatment of uterine and endometrial cancer  
XX and identification of therapeutic agents -  
XX Claim 3; Page 207; 444pp; German.

XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
XX that are highly expressed in uterine tumour tissue and which have  
XX anticancer and cytostatic activity. (A) are used (i) for recombinant  
XX expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
XX are used (i) to identify agents suitable for treatment of uterine or  
XX endometrial cancer; (ii) directly for treating these forms of cancer  
XX (including expression from gene therapy vectors) and (iii) for generation  
XX of specific antibodies. (A) are identified by assembling ESTs (expressed  
XX sequence tags) from a particular tissue type before comparison of the  
XX expression patterns. This allows a significantly longer fragment of the  
XX gene to be revealed, so should reduce the number of failures associated  
XX with the fact that ESTs from different libraries may represent different  
XX parts of the same unknown gene, distorting the estimated frequency of  
XX occurrence in a particular tissue. Z41981-242121 represent EST fragments  
XX derived from a human endometrium tumour cDNA library which encode the  
XX protein sequences represented in Y59941-Y60328.

XX Sequence 1340 BP; 276 A; 374 C; 411 G; 279 T; 0 other;

Query Match 1.6%; Score 17; DB 20; Length 1340;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 tgcctctccctctctcc 38  
|||||  
Db 139 TGCCTCTCCCTCTCTCC 123

RESULT 9  
X13135  
ID X13135 standard; DNA; 1521 BP.  
XX

AC X13135;  
 XX  
 DT 19-MAR-1999 (first entry)  
 XX  
 DE Enterococcus faecalis genome contig SEQ ID NO:198.  
 XX  
 DE Enterococcus faecalis; contig: detection; Enterococcal infection;  
 KW vaccine; attenuation; computer readable medium; ds.  
 KW  
 XX Enterococcus faecalis.  
 OS  
 XX WO9850555-A2.  
 PN 12-NOV-1998.  
 PD  
 XX 04-MAY-1998; 98WO-US08985.  
 PF  
 XX 14-NOV-1997; 97US-0066009.  
 PR 06-MAY-1997; 97US-0044031.  
 PR 16-MAY-1997; 97US-0046655.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Barash SC, Dillon PJ, Kunsch CA;  
 PI WPI; 1999-045171/04.  
 DR  
 XX New isolated Enterococcus faecalis polynucleotides and polypeptides  
 XX - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 XX  
 PS Claim 1; Page 1051-1052; 2084pp; English.  
 XX  
 CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC X12938 to X13919 represent these nucleotide sequences which are primary  
 CC nucleotide sequences, also known as contigs. The computer-based system  
 CC can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 XX  
 SQ Sequence 1521 BP; 458 A; 322 C; 258 G; 480 T; 3 other;

Query Match 1.6%; Score 17; DB 20; Length 1521;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 627 ttggtattcaagtaaat 643  
 |||||  
 Db 394 ttggtattcaagtaaat 410

RESULT 10  
 T30291/C  
 ID T30291 standard; cDNA; 1716 BP.  
 XX  
 AC T30291;  
 XX  
 DT 15-OCT-1996 (first entry)  
 XX  
 DE Human SLAM1 cDNA clone pSURSLAM1.  
 XX  
 KW SLAM1; SLAM1; pSURSLAM1; cell surface antigen; T lymphocyte;  
 KW T-cell co-stimulatory molecule; ss.

XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 61..1068  
 FT /\*tag= a  
 FT sig\_peptide 61..141  
 FT /\*tag= b  
 FT mat\_peptide 142..1065  
 FT /\*tag= c  
 FT exon 761..780  
 FT /\*tag= d  
 FT /\*note= "exon codes for a transmembrane domain  
 FT absent in human SLAM3"  
 XX  
 PN WO9617060-A1.  
 XX  
 PD 06-JUN-1996.  
 XX  
 PF 29-NOV-1995; 95WO-US14942.  
 PR 07-JUN-1995; 95US-0481777.  
 PR 02-DEC-1994; 94US-0348792.  
 XX (SCHE) SCHERING CORP.  
 PA  
 XX Aversa G, Chang C, Cocks BG, De Vries JE;  
 PI WPI; 1996-277780/28.  
 DR P-PSDB; R97628.  
 XX  
 PT New isolated SLAM protein and gene - used to develop prods. for use  
 PT in diagnostics, screening studies and therapy involving  
 PT immunological responses  
 PT  
 XX Claim 10; Page 54-66; 94pp; English.  
 PS  
 XX A cDNA clone (T30291), designated pSURSLAM1 (ATCC 69713), codes for  
 CC human SLAM1 (R97628), a novel T-cell stimulatory molecule which,  
 CC when engaged, potentiates T-cell expansion and induces a Th10/Th1  
 CC cytokine prodn. profile. It was isolated from a T-cell cDNA library  
 CC by expression cloning using anti-SLAM monoclonal antibody A12 for  
 CC selection. Other human SLAM cDNA clones (see also T30292-94) were  
 CC also isolated. pSURSLAM1 was used as a probe to map the SLAM gene  
 CC to a region close to 1q21.3 and 1q22; this region correlates to  
 CC the gene for systemic lupus erythematosus susceptibility. The cDNA  
 CC can be used for prodn. of recombinant SLAM1, or as a probe or primer.  
 XX  
 SQ Sequence 1716 BP; 501 A; 420 C; 395 G; 400 T; 0 other;

Query Match 1.6%; Score 17; DB 17; Length 1716;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 gacagactctggacag 80  
 |||||  
 Db 1014 GACAGACTCTGGACAG 998

RESULT 11  
 Q29117  
 ID Q29117 standard; DNA; 1830 BP.  
 XX  
 AC Q29117;  
 XX  
 DT 24-FEB-1993 (first entry)  
 XX  
 DE HN gene consensus sequence.  
 XX  
 KW Amplify; fusion; F; hemagglutinin neuraminidase; HN; Urabe AM-9;  
 KW mumps virus; polymerase chain reaction; PCR; pBluescript SK+;  
 KW ALVAC; recombinant; entomopox; 42 kD promoter; vaccinia virus;

KW H6 promoter; C3; canarypox virus; virulence factor; deletion loci;  
 KW recipient loci; ss.  
 XX Synthetic.  
 OS  
 XX  
 PN WO9215672-A.  
 XX  
 PD 17-SEP-1992.  
 XX  
 XX 09-MAR-1992; 92WO-US01906.  
 XX  
 XX 07-MAR-1991; 91US-0666056.  
 PR 11-JUN-1991; 91US-0713967.  
 PR 06-MAR-1992; 92US-0847951.  
 XX  
 XX (VIRO-) VIROGENETICS CORP.  
 PA  
 XX Cox WI, De Taisne C, Francis J, Gettig RR, Johnson GP;  
 PI Limbach KJ, Norton EK, Paoletti E, Perkins ME, Pincus SE;  
 PI Riviere M, Tartaglia J, Taylor J;  
 XX WPI; 1992-331718/40.  
 DR  
 XX Vaccine comprises recombinant, attenuated pox-virus - use for  
 PT vaccinating against viral infections such as rabies, hepatitis B,  
 PT HIV, HSV, EBV, CMV, mumps etc.  
 XX  
 PS Disclosure; Fig 30; 456pp; English.  
 XX  
 CC The sequences given in Q29116-17 are consensus sequences of the fusion  
 CC (F) and hemagglutinin neuraminidase (HN) genes from the Urabe AM-9  
 CC strain of mumps virus. They were isolated by polymerase chain reaction  
 CC (PCR). The amplified F and HN genes were digested with SacI and  
 CC Asp718 and cloned into pBluescript SK+. The F and HN genes isolated  
 CC in this manner were used to produce ALVAC recombinants expressing the  
 CC mumps F and HN genes. The F gene was coupled to the entomopox 42 KD  
 CC promoter and the HN gene was coupled to the vaccinia virus H6  
 CC promoter. The two promoted genes were configured in a 5' to 5'  
 CC orientation and inserted into ALVAC at the deleted C3 locus. ALVAC  
 CC is derived from a canarypox virus which has been modified by deletion  
 CC of non-essential regions of the genome encoding known or potential  
 CC virulence factors. The deletion loci were engineered as recipient  
 CC loci for the insertion of foreign genes. See also Q35501-864.  
 XX  
 SQ Sequence 1830 BP; 538 A; 425 C; 344 G; 523 T; 0 other;  
 XX  
 Query Match 1.6%; Score 17; DB 13; Length 1830;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 285 tattgtcactttagggtg 301  
 Db 205 tattgtcactttagggtg 221  
 RESULT 12  
 N92624  
 ID N92624 standard; cDNA; 1900 BP.  
 XX  
 AC N92624;  
 XX  
 XX 14-MAY-1990 (first entry)  
 DT  
 XX Nucleotide sequence encoding mumps virus HN protein.  
 DE  
 XX Haemagglutinin-neuraminidase; mumps; vaccine; ss.  
 KW  
 XX Mumps virus.  
 OS  
 XX  
 XX Kavanaugh WM, Pot D, Williams LT;  
 FH WPI; 1997-341681/31.  
 FT P-PSDB; W26624.  
 FT

XX WO8911534-A.  
 PN  
 XX 30-NOV-1989.  
 PD  
 XX 17-MAY-1989; 89WO-US02130.  
 PF  
 XX 24-MAY-1988; 88US-0198451.  
 PR  
 XX (TEXA ) UNIV OF TEXAS SYSTE.  
 PA  
 XX Waxham MN;  
 PI  
 XX WPI; 1989-370727/50.  
 DR  
 XX P-PSDB; P93666  
 DR  
 XX HN specific degenerate oligonucleotide probe - useful for screening host  
 PT cells producing HN to use prod. in vaccine development  
 FT  
 PS Claim 1; Page 25; 40pp; English.  
 PS  
 XX Nucleotide sequence encoding HN protein of mumps virus, which may be used  
 CC to generate a large amounts of purified peptides from a unicellular host.  
 CC  
 XX Sequence 1900 BP; 569 A; 433 C; 365 G; 533 T; 0 other;  
 SQ  
 Query Match 1.6%; Score 17; DB 10; Length 1900;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 285 tattgtcactttagggtg 301  
 Db 234 tattgtcactttagggtg 250  
 RESULT 13  
 T90501  
 ID T90501 standard; cDNA; 2940 BP.  
 XX  
 AC T90501;  
 XX  
 XX 28-JAN-1998 (first entry)  
 DT  
 XX Signalling inositol polyphosphate 5-phosphatase SIP-110 cDNA.  
 DE  
 XX SIP-110; signalling inositol polyphosphate 5-phosphatase; human;  
 KW signal transduction; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH CDS 10..2940  
 FT /\*tag= a  
 FT  
 XX WO9722690-A2.  
 PN  
 XX 26-JUN-1997.  
 PD  
 XX 06-DEC-1996; 96WO-US19515.  
 PF  
 XX 04-DEC-1996; 96US-0759397.  
 PR 08-DEC-1995; 95US-0569378.  
 PR 14-DEC-1995; 95US-0008607.  
 PR 28-MAR-1996; 96US-0624190.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA  
 XX Kavanaugh WM, Pot D, Williams LT;  
 PI  
 XX WPI; 1997-341681/31.  
 DR P-PSDB; W26624.  
 DR  
 XX

PT Signalling inositol polyphosphate 5-phosphatase, SIP 130 and  
 PT related DNA - used for treating abnormal cell growth, regulating  
 PT mitogenic activity and calcium signalling  
 XX  
 PS Disclosure; Page 78-79; 114pp; English.

XX This cDNA sequence encodes a signalling inositol polyphosphate  
 CC 5-phosphatase (SIP), designated SIP-110 (see W26624), also known  
 CC as GRB2-associated inositol polyphosphate 5-phosphatase. SIP-110  
 CC binds both SH3 domains on the GRB2 protein, and modulates  
 CC signalling of e.g. ras. SIP-110 is a splice variant of novel  
 CC SIP-130 (see W26623). SIP-130 polyptides and polynucleotides can  
 CC be used in claimed methods for modulating phosphatidylinositol  
 CC 3-kinase activity, mitogen-activated protein (MAP) kinase activity,  
 CC the level of phosphatidylinositol (3,4,5) triphosphate in a cell,  
 CC allowing treatment of diseases associated with abnormal cell  
 CC growth.

XX Sequence 2940 BP; 730 A; 903 C; 778 G; 529 T; 0 other;

Query Match 1.6%; Score 17; DB 18; Length 2940;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 299 gtgggaagtcactagca 315  
 Db 2046 gtgggaagtcactagca 2062  
 |||||

## RESULT 14

249226/c  
 ID 249226 standard; cDNA; 3203 BP.

XX 249226;  
 AC

XX 07-MAR-2000 (first entry)

XX Human hydrolase homologue HHH-4 cDNA.

DE Hydrolase; homologue; HHH-4; N-terminal asparagine amidohydrolase;  
 KW vanin-1; glycosyl hydrolase; glucosyl hydrolase;  
 KW N-acetylglucosamine 6-P deacetylase; diagnosis;  
 KW treatment; prevention; expression; disorder; carbohydrate; metabolism;  
 KW antagonist; reproductive disorder; cell proliferation; ds.

XX Synthetic.

OS Homo sapiens.

XX Key Location/Qualifiers  
 CDS 162..1367

FT //tag= a

FT /product "HHH-4"

FT /note= "Homology with glycosyl hydrolase (BL00775)"

XX W09961626-A2.

XX 02-DEC-1999.

XX 28-MAY-1999; 99WO-US12021.

XX 29-MAY-1998; 98US-0087236.

XX (INCY-) INCYTE PHARM INC.

XX Bandman O, Hillman JL, Yue H, Lal P, Corley NC, Guegler KJ;

PI Patterson C, Baughn MR;

XX WPI; 2000-062716/05.

DR P-PSDB; Y58168.

XX New human hydrolase homologues, useful for treating or preventing a  
 PT carbohydrate metabolism disorder

XX Claim 7; Page 63-70; 91pp; English.

XX This sequence represents cDNA encoding a human hydrolase homologue,  
 CC HHH-4. Nucleic acids encoding HHH-4 were initially identified in a  
 CC neonatal dermal microvascular endothelial cell cDNA library and the  
 CC cDNA sequence includes the cDNA fragments 249232, 249234, 249236,  
 CC 249238, 249241, 249243, 249244, 249245, 249246, 249253, 249257, and  
 CC 249261. HHH-4 has homology to glycosyl hydrolase (BL00775). Glycosyl  
 CC hydrolases cleave the glycosidic bond between two carbohydrates or  
 CC between a carbohydrate and a non-carbohydrate moiety, and are frequently  
 CC associated with disorders of carbohydrate metabolism. The invention  
 CC relates to human hydrolase homologues HHH-1 to HHH-7 (Y58165-Y58171)  
 CC which respectively have homology to N-terminal asparagine amidohydrolase,  
 CC vanin-1, glycosyl hydrolases, glucosyl hydrolase and N-acetylglucosamine 6-P  
 CC deacetylase. Such homologues are useful in methods for diagnosing,  
 CC treating or preventing disorders associated with expression of  
 CC hydrolases. The hydrolase homologues are useful for treating or  
 CC preventing a carbohydrate metabolism disorder. Antagonists of these  
 CC hydrolases can be used to treat or prevent a reproductive or cell  
 CC proliferation disorder.

XX Sequence 3203 BP; 562 A; 903 C; 887 G; 751 T; 0 other;

Query Match 1.6%; Score 17; DB 21; Length 3203;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 tgctctccctctctcc 38  
 |||||

Db 654 TGCCTCTCCCTTCTCC 638

## RESULT 15

T90500

ID T90500 standard; cDNA; 3691 BP.

XX T90500;

XX 28-JAN-1998 (first entry)

XX Signalling inositol polyphosphate 5-phosphatase SIP-130 cDNA.

XX SIP-130; signalling inositol polyphosphate 5-phosphatase; human;

KW signal transduction; phosphatidylinositol 3'-kinase;

KW mitogen activated protein kinase; MAP kinase;

KW phosphatidylinositol (3,4,5) triphosphate; calcium signalling;

KW mitogen; cell growth; cell proliferation; apoptosis; gene delivery;

gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 CDS 122..3691

FT //tag= a

XX W09722690-A2.

XX 26-JUN-1997.

XX 06-DEC-1996; 96WO-US19515.

XX 04-DEC-1996; 96US-0759397.

XX 08-DEC-1995; 95US-0569578.

XX 14-DEC-1995; 95US-0008607.

XX 28-MAR-1996; 96US-0624190.

XX (CHIR ) CHIRON CORP.

XX Kavanaugh WM, Pot D, Williams LT;

XX WPI; 1997-341681/31.



us-09-101-423a-1.olig.rng

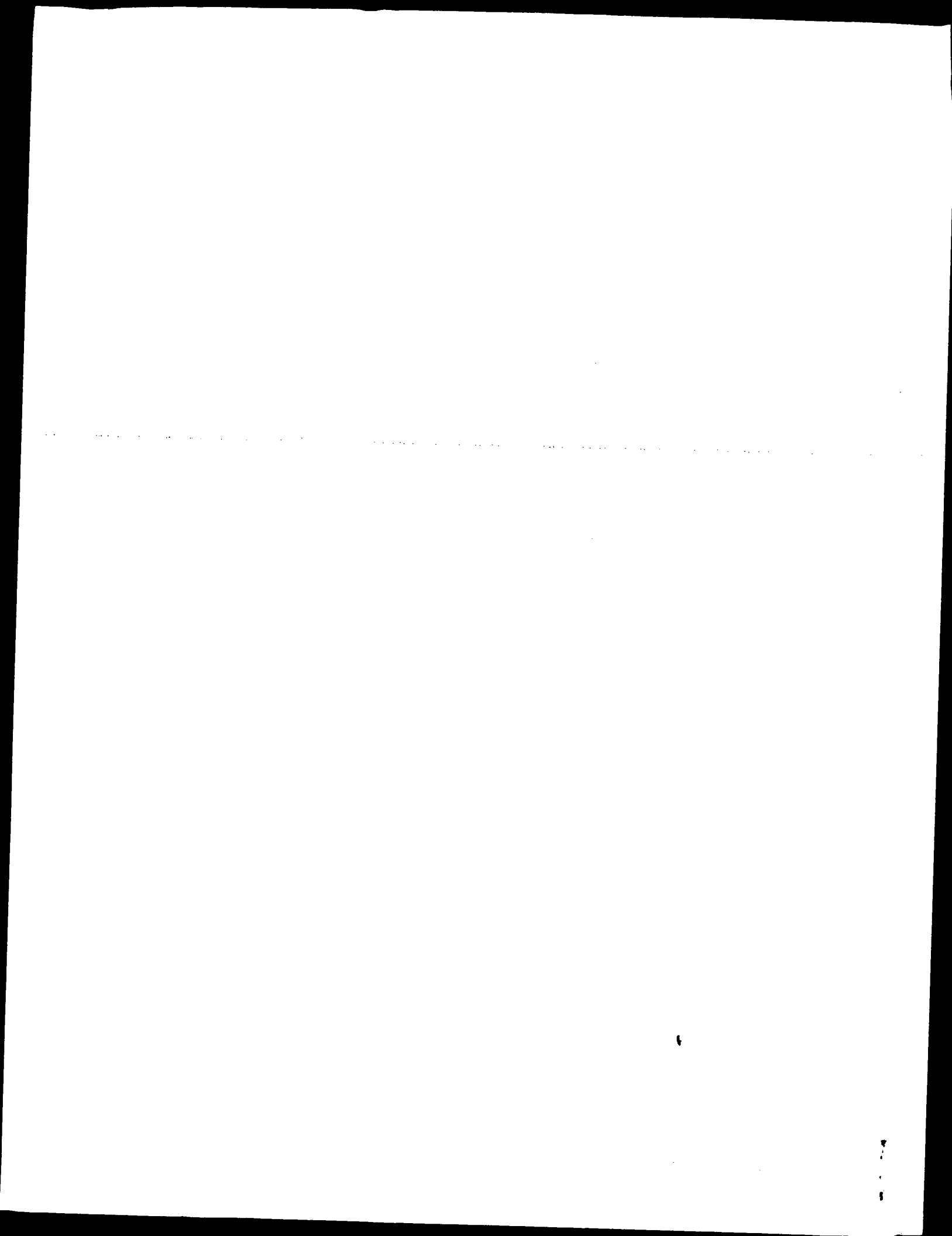
- Mon Nov 6 10:14:01 2000

DR P-PSDB; W26623.  
 XX Signalling inositol polyphosphate 5-phosphatase, SIP 130 and  
 PT related DNA - used for treating abnormal cell growth, regulating  
 PT mitogenic activity and calcium signalling  
 XX  
 XX  
 PS Claim 7; Page 80-82; 114pp; English.  
 XX  
 CC This cDNA clone encodes a novel signalling inositol polyphosphate  
 CC 5-phosphatase (SIP), designated SIP-130 (see W26623), capable of  
 CC binding to a phosphotyrosine binding domain of SH2 and collagen  
 CC containing protein (SHC). It was isolated from a human lung cDNA  
 CC library using probes based in SIP-110 sequences. SIP-110 (see  
 CC T90501) and SIP-N (see T90502) were identified as splice variants  
 CC of SIP-130. SIP-130 nucleic acids can be used to produce SIP-130  
 CC polypeptides in transformed host cells. Coding or non-coding  
 CC sequences (including 5' or 3' untranslated regions (UTR), a  
 CC promoter region, or an antisense or ribozyme sequence) can be used  
 CC to modulate SIP activity. SIP mRNA transcription or the level of SIP  
 CC protein expression (claimed). Methods are claimed for modulating  
 CC phosphatidylinositol 3-kinase activity, mitogen-activated protein  
 CC (MAP) kinase activity, the level of phosphatidylinositol (3,4,5)  
 CC triphosphate in a cell, or mitogenic activity using SIP or  
 CC SIP-modulating polynucleotides, allowing treatment of diseases  
 CC associated with abnormal cell growth and calcium signalling. The  
 CC UTRs and promoter regions can also be used to regulate a  
 CC heterologous coding sequence.  
 XX  
 SQ Sequence 3691 BP; 906 A; 1125 C; 997 G; 663 T; 0 other;

Query Match 1.6%; Score 17; DB 18; Length 3691;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 299 gtgggaagtactagca 315  
 |||||  
 Db 2797 gtgggaagtactagca 2813

Search completed: November 4, 2000, 03:35:23  
 Job time: 10987 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:13:30 ; Search time 140.17 Seconds  
(without alignments)  
1114.592 Million cell updates/sec

Title: US-09-101-423A-1  
Perfect score: 1033  
Sequence: 1 ctctcttggtctatgtc.....gtacacataactccatgggt 1033

Scoring table: OLIGO\_NUC  
Gapopen 60.0 , Gapext 60.0

Searched: 262060 seqs, 75620727 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/PCITUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17	1.6	1020	1	US-08-348-792-5
C 2	17	1.6	1020	4	US-08-462-738-5
C 3	17	1.6	1079	1	US-08-348-792-7
C 4	17	1.6	1079	4	US-08-462-738-7
C 5	17	1.6	1716	1	US-08-348-792-1
C 6	17	1.6	1716	4	US-08-462-738-1
C 7	17	1.6	1830	1	US-08-105-483-371
C 8	17	1.6	1830	2	US-08-709-209-371
C 9	17	1.6	1830	2	US-08-458-101-371
C 10	17	1.6	2940	5	US-09-195-868-11
C 11	17	1.6	3691	5	US-09-195-868-12
C 12	17	1.6	4147	5	US-08-560-005-1
C 13	16	1.5	20	5	US-08-922-505A-35
C 14	16	1.5	349	1	US-08-543-238-3
C 15	16	1.5	349	1	US-08-420-526-3
C 16	16	1.5	437	1	US-08-543-238-1
C 17	16	1.5	437	1	US-08-420-526-1
C 18	16	1.5	1000	4	US-07-718-538-1
C 19	16	1.5	1245	1	US-07-887-072B-1
C 20	16	1.5	1245	2	US-08-466-444-1
C 21	16	1.5	1650	3	US-08-436-771-10
C 22	16	1.5	1650	3	US-08-434-988-10
C 23	16	1.5	1650	3	US-08-487-797-10
C 24	16	1.5	1650	6	PCT-US95-02058-10
C 25	16	1.5	1767	3	US-08-841-178-21
C 26	16	1.5	1921	3	US-08-841-178-23

c	27	16	1.5	2270	3	US-08-852-807-16	Sequence 16, Appl
	28	16	1.5	2990	1	US-07-671-817A-1	Sequence 1, Appl
	29	16	1.5	3531	1	US-07-828-788A-7	Sequence 7, Appl
	30	16	1.5	3531	1	US-07-920-085-1	Sequence 1, Appl
	31	16	1.5	3531	5	US-08-754-490-11	Sequence 11, Appl
	32	16	1.5	3531	5	US-08-922-505A-11	Sequence 11, Appl
	33	16	1.5	3531	6	PCT-US92-11337-7	Sequence 7, Appl
	34	16	1.5	3531	7	5169629-1	Patent No. 5169629
	35	16	1.5	3534	3	US-08-841-178-24	Sequence 24, Appl
	36	16	1.5	3534	3	US-08-841-178-25	Sequence 25, Appl
	37	16	1.5	3534	3	US-08-841-178-26	Sequence 26, Appl
	38	16	1.5	3534	5	US-08-922-505A-33	Sequence 33, Appl
	39	16	1.5	3537	1	US-08-377-690-3	Sequence 3, Appl
	40	16	1.5	3537	5	US-08-855-160-1	Sequence 1, Appl
c	41	16	1.5	13674	3	US-08-852-807-1	Sequence 1, Appl
	42	16	1.5	51259	5	US-08-781-891-209	Sequence 209, App
	43	15	1.5	27	4	US-08-859-998-905	Sequence 905, App
c	44	15	1.5	28	5	US-08-754-490-3	Sequence 3, Appl
c	45	15	1.5	28	5	US-08-922-505A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-348-792-5/c  
; Sequence 5, Application US/08348792  
; Patent No. 5376423  
; GENERAL INFORMATION:  
; APPLICANT: Aversa, Gregorio  
; APPLICANT: Chang, Chia-Chun J.  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: de Vries, Jan E.  
; TITLE OF INVENTION: PURIFIED GENES ENCODING MAMMALIAN CELL  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: INAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/348,792  
; FILING DATE: 02-DEC-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0436  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-852-9196  
; TELEFAX: 415-496-1200  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1020 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 61...975  
; US-08-348-792-5

Query Match 1.6%; Score 17; DB 1; Length 1020;

Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 gacagactctgggacag 80  
|||||  
Db 924 GACAGACTCTGGGACAG 908

## RESULT 2

US-08-462-738-5/c  
; Sequence 5, Application US/08462738  
; Patent No. 5977303  
; GENERAL INFORMATION:  
; APPLICANT: Aversa, Gregorio  
; APPLICANT: Chang, Chia-Chun J.  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: de Vries, Jan E.  
; TITLE OF INVENTION: PURIFIED GENES ENCODING MAMMALIAN CELL  
; TITLE OF INVENTION: SURFACE ANTIGENS; PROTEINS AND ANTIBODIES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,738  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/348,792  
; FILING DATE: 02-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-852-9196  
; TELEFAX: 415-496-1200  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1020 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 61..975  
US-08-462-738-5

Query Match 1.6%; Score 17; DB 4; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 gacagactctgggacag 80  
|||||  
Db 924 GACAGACTCTGGGACAG 908

## RESULT 3

US-08-348-792-7/c  
; Sequence 7, Application US/08348792  
; Patent No. 5576423  
; GENERAL INFORMATION:

; APPLICANT: Aversa, Gregorio  
; APPLICANT: Chang, Chia-Chun J.  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: de Vries, Jan E.  
; TITLE OF INVENTION: PURIFIED GENES ENCODING MAMMALIAN CELL  
; TITLE OF INVENTION: SURFACE ANTIGENS; PROTEINS AND ANTIBODIES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/348,792  
; FILING DATE: 02-DEC-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0436  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-852-9196  
; TELEFAX: 415-496-1200  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1079 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 153..1073  
US-08-348-792-7

Query Match 1.6%; Score 17; DB 1; Length 1079;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 gacagactctgggacag 80  
|||||  
Db 1022 GACAGACTCTGGGACAG 1006

## RESULT 4

US-08-462-738-7/c  
; Sequence 7, Application US/08462738  
; Patent No. 5977303  
; GENERAL INFORMATION:  
; APPLICANT: Aversa, Gregorio  
; APPLICANT: Chang, Chia-Chun J.  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: de Vries, Jan E.  
; TITLE OF INVENTION: PURIFIED GENES ENCODING MAMMALIAN CELL  
; TITLE OF INVENTION: SURFACE ANTIGENS; PROTEINS AND ANTIBODIES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,738  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,792  
FILING DATE: 02-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0436GB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1079 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 153..1073  
US-08-462-738-7

Query Match 1.6% Score 17; DB 4; Length 1079;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 gacagactctgggacag 80  
|||||

Db 1022 GACAGACTCTGGGACAG 1006

RESULT 5  
US-08-348-792-1/c  
Sequence 1, Application US/08348792  
Patent No. 5576423  
GENERAL INFORMATION:  
APPLICANT: Aversa, Gregorio  
APPLICANT: Chang, Chia-Chun J.  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: de Vries, Jan E.  
TITLE OF INVENTION: PURIFIED GENES ENCODING MAMMALIAN CELL  
SURFACE ANTIGENS; PROTEINS AND ANTIBODIES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/348,792  
FILING DATE: 02-DEC-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0436  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196

TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1716 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 61..1065  
US-08-348-792-1

Query Match 1.6% Score 17; DB 1; Length 1716;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 gacagactctgggacag 80  
|||||

Db 1014 GACAGACTCTGGGACAG 998

RESULT 6  
US-08-462-738-1/c  
Sequence 1, Application US/08462738  
Patent No. 5977303  
GENERAL INFORMATION:  
APPLICANT: Aversa, Gregorio  
APPLICANT: Chang, Chia-Chun J.  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: de Vries, Jan E.  
TITLE OF INVENTION: PURIFIED GENES ENCODING MAMMALIAN CELL  
SURFACE ANTIGENS; PROTEINS AND ANTIBODIES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,738  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,792  
FILING DATE: 02-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0436GB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-852-9196  
TELEFAX: 415-496-1200  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1716 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 61..1065  
US-08-462-738-1

Query Match 1.6%; Score 17; DB 4; Length 1716;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 gacagactctgggacag 80  
|||||  
DB 1014 GACAGACTCTGGGACAG 998

## RESULT 7

US-08-105-483-371  
; Sequence 371, Application US/08105483  
; Patent No. 5494807  
; GENERAL INFORMATION:  
; APPLICANT: Paolletti, Enzo  
; TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE  
; NUMBER OF SEQUENCES: 462  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; ADDRESSEE: c/o William S. Frommer  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/105,483  
; FILING DATE: 12-AUG-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/847,951  
; FILING DATE: 06-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer, William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454310-2400  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; INFORMATION FOR SEQ ID NO: 371:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1830 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-105-483-371

Query Match 1.6%; Score 17; DB 1; Length 1830;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 tattgtcacttttagtg 301  
|||||  
DB 205 TATTGTCACCTTAGGTG 221

## RESULT 8

US-08-709-209-371  
; Sequence 371, Application US/08709209  
; Patent No. 5762938  
; GENERAL INFORMATION:  
; APPLICANT: Paolletti, Enzo  
; TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE  
; NUMBER OF SEQUENCES: 462

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; ADDRESSEE: c/o William S. Frommer  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/709,209  
; FILING DATE: 21-AUG-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/105,483  
; FILING DATE: 12-AUG-1993  
; APPLICATION NUMBER: US 07/847,951  
; FILING DATE: 06-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer, William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454310-2400  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; INFORMATION FOR SEQ ID NO: 371:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1830 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-709-209-371

Query Match 1.6%; Score 17; DB 2; Length 1830;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 tattgtcacttttagtg 301  
|||||  
DB 205 TATTGTCACCTTAGGTG 221

## RESULT 9

US-08-458-101-371  
; Sequence 371, Application US/08458101  
; Patent No. 5766599  
; GENERAL INFORMATION:  
; APPLICANT: Paolletti, Enzo  
; APPLICANT: Perkus, Marion E.  
; APPLICANT: Taylor, Jill  
; APPLICANT: Taraglia, James  
; APPLICANT: No. 5766599ton, Elizabeth K.  
; APPLICANT: Riviere, Michel  
; APPLICANT: de Gaisne, Charles  
; APPLICANT: Limbach, Keith J.  
; APPLICANT: Johnson, Gerard P.  
; APPLICANT: Pincus, Steven E.  
; APPLICANT: Cox, William I.  
; APPLICANT: Audonnet, Jean-Christophe Francis  
; APPLICANT: Gettig, Russell Robert  
; TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE  
; TITLE OF INVENTION: STRAIN  
; NUMBER OF SEQUENCES: 467  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; ADDRESSEE: c/o William S. Frommer  
; STREET: 530 Fifth Avenue  
; CITY: New York

```

; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,101
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2740
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 371:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-458-101-371
;
; Query Match 1.6%; Score 17; DB 2; Length 1830;
; Best Local Similarity 100.0%; Pred. No. 15;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 285 tattgtcacttcagatg 301
Db 205 TATTGTCACTTAGGTG 221
;
; RESULT 10
; US-09-195-868-11
; Sequence 11, Application US/09195868
; Patent No. 6050621
; GENERAL INFORMATION:
; APPLICANT: KAVANAUGH MD, MICHAEL
; APPLICANT: POT PH.D., DAVID
; APPLICANT: WILLIAMS MDPHD, LEWIS T.
; TITLE OF INVENTION: SIGNALING INOSITOL POLYPHOSPHATE
; TITLE OF INVENTION: 5-PHOSPHATASES (SIPs)
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 HORTON STREET
; CITY: EMERYVILLE
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; CLASSIFICATION:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FIRESTONE, LEIGH H.
; REGISTRATION NUMBER: 36,831
; REFERENCE/DOCKET NUMBER: 1182.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-923-2707
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3691 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-195-868-12
;
; Query Match 1.6%; Score 17; DB 5; Length 3691;
; Best Local Similarity 100.0%; Pred. No. 16;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 299 gtgggaagtcactagca 315
Db 2797 GTGGGAAGTCACTAGCA 2813
;
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2940 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-195-868-11
;
; Query Match 1.6%; Score 17; DB 5; Length 2940;
; Best Local Similarity 100.0%; Pred. No. 15;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 299 gtgggaagtcactagca 315
Db 2046 GTGGGAAGTCACTAGCA 2062
;
; RESULT 11
; US-09-195-868-12
; Sequence 12, Application US/09195868
; Patent No. 6050621
; GENERAL INFORMATION:
; APPLICANT: KAVANAUGH MD, MICHAEL
; APPLICANT: POT PH.D., DAVID
; APPLICANT: WILLIAMS MDPHD, LEWIS T.
; TITLE OF INVENTION: SIGNALING INOSITOL POLYPHOSPHATE
; TITLE OF INVENTION: 5-PHOSPHATASES (SIPs)
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 HORTON STREET
; CITY: EMERYVILLE
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; CLASSIFICATION:
; APPLICATION NUMBER: US/09/195,868
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FIRESTONE, LEIGH H.
; REGISTRATION NUMBER: 36,831
; REFERENCE/DOCKET NUMBER: 1182.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-923-2707
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3691 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-195-868-12
;
; Query Match 1.6%; Score 17; DB 5; Length 3691;
; Best Local Similarity 100.0%; Pred. No. 16;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 299 gtgggaagtcactagca 315
Db 2797 GTGGGAAGTCACTAGCA 2813
;

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

Query Match      1.6%; Score 17; DB 5; Length 4147;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      299  gtgggaagtcactagca 315
          |||
Db      2053  GTGGGAAGTCACTAGCA 2069

RESULT 13
US-08-922-505A-35
Sequence 35 Application US/08922505A
Patent No. 6110464
GENERAL INFORMATION:
APPLICANT: Malvar, Thomas
APPLICANT: Gilmer, Amy Jelen
TITLE OF INVENTION: BROAD-SPECTRUM (-ENDOTOXINS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA

```



STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Beta vulgaris  
US-08-543-238-3

Query Match 1.5%; Score 16; DB 1; Length 349;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 tcacatctgagaaagg 342  
|||||

Db 287 TCACATCTGAGAAAGG 272

## RESULT 15

US-08-420-526-3/c

; Sequence 3, Application US/08420526

; Patent No. 5608151

; GENERAL INFORMATION:

; APPLICANT: Bojsen, Kirsten

; APPLICANT: Kragh, Karsten M.

; APPLICANT: Mikkelsen, Jorn D.

; APPLICANT: Nielsen, Klaus K.

; TITLE OF INVENTION: Anti-Microbial Proteins

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sandoz Agro, Inc.

; STREET: 975 California Avenue

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/420,526

; FILING DATE:

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Marcus-Wyner, Lynn

; REGISTRATION NUMBER: 34,869

; REFERENCE/DOCKET NUMBER: 137-1078/WA

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/354-3588

; TELEFAX: 415/857-1125

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 349 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: unknown

; MOLECULE TYPE: CDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Beta vulgaris

; US-08-420-526-3

Query Match 1.5%; Score 16; DB 1; Length 349;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 tcacatctgagaaagg 342  
|||||

Db 287 TCACATCTGAGAAAGG 272

Search completed: November 4, 2000, 03:30:55  
Job time: 11845 sec

Mon Nov 6 10:14:02 2000

us-09-101-423a-1.olig.rni

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2000, 20:18:19 ; Search time 3299.68 Seconds  
(without alignments)  
1935.593 Million cell updates/sec

Title: US-09-101-423A-1  
Perfect score: 1033  
Sequence: 1 ctctctgggtgctctatgtc.....gtagacataactcccatggt 1033

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 7189864 seqs, 3091403243 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*
- 10: gb\_est10.\*
- 11: gb\_est11.\*
- 12: gb\_est12.\*
- 13: gb\_est13.\*
- 14: gb\_est14.\*
- 15: gb\_est15.\*
- 16: gb\_est16.\*
- 17: gb\_est17.\*
- 18: gb\_est18.\*
- 19: gb\_est19.\*
- 20: gb\_est20.\*
- 21: gb\_est21.\*
- 22: gb\_est22.\*
- 23: gb\_est23.\*
- 24: gb\_est24.\*
- 25: gb\_est25.\*
- 26: gb\_est26.\*
- 27: gb\_est27.\*
- 28: gb\_est28.\*
- 29: gb\_est29.\*
- 30: gb\_est30.\*
- 31: gb\_est31.\*
- 32: gb\_est32.\*
- 33: gb\_est33.\*
- 34: gb\_est34.\*
- 35: gb\_est35.\*
- 36: gb\_est36.\*
- 37: gb\_est37.\*
- 38: gb\_est38.\*
- 39: gb\_est39.\*
- 40: gb\_est40.\*
- 41: em\_estba.\*
- 42: em\_estfun.\*
- 43: em\_esthum1.\*

- 44: em\_esthum2.\*
- 45: em\_esthum3.\*
- 46: em\_esthum4.\*
- 47: em\_esthum5.\*
- 48: em\_esthum6.\*
- 49: em\_esthum7.\*
- 50: em\_esthum8.\*
- 51: em\_esthum9.\*
- 52: em\_esthum10.\*
- 53: em\_esthum11.\*
- 54: em\_esthum12.\*
- 55: em\_esthum13.\*
- 56: em\_esthum14.\*
- 57: em\_esthum15.\*
- 58: em\_esthum16.\*
- 59: em\_esthum17.\*
- 60: em\_esthum18.\*
- 61: em\_esthum19.\*
- 62: em\_esthum20.\*
- 63: em\_estin1.\*
- 64: em\_estin2.\*
- 65: em\_estin3.\*
- 66: em\_estin4.\*
- 67: em\_estom.\*
- 68: em\_estov1.\*
- 69: em\_estov2.\*
- 70: em\_estpl1.\*
- 71: em\_estpl2.\*
- 72: em\_estpl3.\*
- 73: em\_estpl4.\*
- 74: em\_estpl5.\*
- 75: em\_estpl6.\*
- 76: em\_estro1.\*
- 77: em\_estro2.\*
- 78: em\_estro3.\*
- 79: em\_estro4.\*
- 80: em\_estro5.\*
- 81: em\_estro6.\*
- 82: em\_estro7.\*
- 83: em\_estro8.\*
- 84: em\_estro9.\*
- 85: em\_estro10.\*
- 86: em\_estro11.\*
- 87: em\_estro12.\*
- 88: em\_estro13.\*
- 89: gb\_gss1.\*
- 90: gb\_gss2.\*
- 91: gb\_gss3.\*
- 92: gb\_gss4.\*
- 93: em\_gss1.\*
- 94: em\_gss2.\*
- 95: em\_gss3.\*
- 96: em\_gss4.\*
- 97: gb\_gss5.\*
- 98: gb\_gss6.\*
- 99: gb\_gss7.\*
- 100: gb\_gss8.\*
- 101: gb\_gss9.\*
- 102: em\_gss5.\*
- 103: em\_gss6.\*
- 104: em\_gss7.\*
- 105: em\_gss8.\*
- 106: em\_gss9.\*
- 107: em\_gss10.\*
- 108: em\_gss11.\*
- 109: gb\_gss10.\*
- 110: gb\_gss11.\*
- 111: em\_gss12.\*
- 112: gb\_gss12.\*
- 113: gb\_gss13.\*
- 114: gb\_gss14.\*
- 115: gb\_gss15.\*
- 116: gb\_gss16.\*
- 117: gb\_gss17.\*

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

117: gb\_gss18:\*  
118: gb\_gss19:\*  
119: em\_gss13:\*  
120: gb\_gss20:\*  
121: gb\_gss21:\*  
122: gb\_gss22:\*  
123: gb\_gss23:\*  
124: gb\_gss24:\*  
125: em\_gss14:\*  
126: em\_gss15:\*  
127: em\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	21	2.0	424	100	A0618877	A0618877 HS_5172_B
C 2	21	2.0	520	117	A2336681	A2336681 RPCI-23-8
C 3	20	1.9	109	12	A1740659	A1740659 wg07a10.x
C 4	20	1.9	230	15	AV060214	AV060214 AV060214
C 5	20	1.9	272	15	AV062192	AV062192 AV062192
C 6	20	1.9	295	91	AQ300268	AQ300268 HS_3170_B
C 7	20	1.9	372	32	BE003749	BE003749 QV3-BN009
C 8	20	1.9	421	40	W42753	W42753 zc24g06.s1
C 9	20	1.9	594	91	AQ315018	AQ315018 RPCI11-94
C 10	20	1.9	621	113	A2015292	A2015292 RPCI-23-2
C 11	20	1.9	625	21	AW360531	AW360531 687053F01
C 12	19	1.8	259	89	AQ106028	AQ106028 HS_3054_A
C 13	19	1.8	263	33	BE142685	BE142685 ILO-HT015
C 14	19	1.8	306	36	D18871	D18871 MUSGS01053
C 15	19	1.8	410	10	A1466102	A1466102 vw40h11.y
C 16	19	1.8	417	108	A0684969	A0684969 HS_2147_A
C 17	19	1.8	423	6	AA823123	AA823123 vw40h11.x
C 18	19	1.8	429	113	AQ981050	AQ981050 RPCI-23-3
C 19	19	1.8	435	111	AQ825108	AQ825108 HS_3249_A
C 20	19	1.8	520	98	AQ5111273	AQ5111273 HS_5193_A
C 21	19	1.8	540	108	AQ695102	AQ695102 HS_5485_B
C 22	19	1.8	560	23	AW660674	AW660674 99654 MAR
C 23	19	1.8	659	35	BE376233	BE376233 601328594
C 24	19	1.8	709	15	AU085759	AU085759 AU085759
C 25	19	1.8	802	122	CNS02XW	AL218993 Tetraodon
C 26	19	1.8	834	118	BI9186	BI9186 F3D15-T7 IG
C 27	19	1.8	930	123	CNS03YPE	AL266503 Tetraodon
C 28	19	1.8	945	123	CNS0480N	AL278672 Tetraodon
C 29	19	1.8	1036	123	CNS0486H	AL304802 Tetraodon
C 30	19	1.8	1112	118	BI0172	BI0172 F15A2-T7 IG
C 31	18	1.7	113	113	AZ000895	AZ000895 RPCI-23-2
C 32	18	1.7	144	20	AW122167	AW122167 UI-M-BH2.
C 33	18	1.7	166	14	AJ271012	AJ271012 AJ271012
C 34	18	1.7	215	10	AI381747	AI381747 te42f09.x
C 35	18	1.7	215	23	AW588076	AW588076 kp18401.y
C 36	18	1.7	220	13	AI848046	AI848046 UI-M-AH1.
C 37	18	1.7	225	18	AV379636	AV379636 AV379636
C 38	18	1.7	241	10	AI462507	AI462507 ub73f04.x
C 39	18	1.7	248	16	AV122459	AV122459 AV122459
C 40	18	1.7	263	7	AA873701	AA873701 ob36h03.s
C 41	18	1.7	263	22	AW445329	AW445329 81383 MAR
C 42	18	1.7	268	31	BB465986	BB465986 BB465986
C 43	18	1.7	276	15	AV061909	AV061909 AV061909
C 44	18	1.7	278	16	AV158405	AV158405 AV158405
C 45	18	1.7	279	16	AV163182	AV163182 AV163182

ALIGNMENTS

RESULT 1  
A0618877/c

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

117: gb\_gss18:\*  
118: gb\_gss19:\*  
119: em\_gss13:\*  
120: gb\_gss20:\*  
121: gb\_gss21:\*  
122: gb\_gss22:\*  
123: gb\_gss23:\*  
124: gb\_gss24:\*  
125: em\_gss14:\*  
126: em\_gss15:\*  
127: em\_gss16:\*

us-09-101-423a-1.olig.rst

Page 2

## AUTHORS

Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret  
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.  
and Fraser,C.M.

## TITLE

Mouse BAC End Sequences from Library RPCI-23

## JOURNAL

Unpublished (1999)

## COMMENT

Other\_GSSs: RPCI-23-87D8.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 87 row: D column: 8  
Seq primer: T7  
Class: BAC ends.

## FEATURES

## source

Location/Qualifiers

1. 520  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-87D8"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site.1:  
EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 156 a 115 c 113 g 136 t  
ORIGIN

Query Match 2.0%; Score 21; DB 117; Length 520;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 693 tgcagctgttcctgtgattg 713

Db 294 TGCAGCTTGTCTGTGATTG 274

## RESULT 3

AI740659

LOCUS

DEFINITION

w07a10.x1 Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone  
IMAGE:2364378.3, similar to SW:TDXN\_HUMAN Q13162 ANTIOXIDANT ENZYME

ACE372 ; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 109)

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.  
Trace considered overall poor quality  
Insert length: 495 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 1.

## FEATURES

## source

Location/Qualifiers  
1. 109  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2364378"  
/clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site.1: Not I; Site.2: Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in  
a subtractive hybridization reaction. The driver was  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares NBHSF pool 1:  
309384-310919, 323208-325895 Soares NB2HP pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares NB2HF8-9W pool 1:  
758280-760583, 772104-774407 Soares NBHPA pool 1:  
304776-306311, 320136-322823, 326280-326563 Soares NBHOT  
pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."

pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 30 a 21 c 24 g 34 t  
ORIGIN

Query Match 1.9%; Score 20; DB 12; Length 109;  
Best Local Similarity 100.0%; Pred. No. 7.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 317 atctgatgggtcacatctga 336

Db 9 ATCTGATGGGTCAACATCTGA 28

## RESULT 4

AV060214/c

LOCUS

DEFINITION

clone 1810061019, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1. (bases 1 to 230)

REFERENCE

AUTHORS

Carinci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,  
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara  
A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawal,J.,  
Kikuchi,N., Kojima,Y., Matsuyama,T., Niituma,H., Oda,H., Owa,C.,  
Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara  
Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomimaga,N.,  
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,  
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

RIKEN Mouse ESTs

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@cc.riken.go.jp

Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

## FEATURES

source

Location/Qualifiers  
1. .230  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="1810061019"  
/clone\_lib="Mus musculus pancreas C57BL/6J adult"  
/sex="male"  
/tissue\_type="pancreas"  
/dev\_stage="adult"  
62 a 58 c 40 g 70 t

BASE COUNT

ORIGIN

Query Match 1.9%; Score 20; DB 15; Length 230;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 214 taaagtgaagtctggggagcc 233

Db 89 TAAAGTGAGTCTGGGGAGCC 70  
|||||

RESULT 5

AV062192/c

LOCUS

DEFINITION AV062192 Mus musculus small intestine C57BL/6J adult Mus musculus  
cDNA clone 2010002H22, mRNA sequence.

ACCESSION

AV062192

VERSION

AV062192.1

KEYWORDS

EST.

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 272)

AUTHORS

Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,  
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara  
A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,  
Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsu,H., Oda,H., Owa,C.,  
Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara  
Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomioka,N.,  
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,  
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

RIKEN Mouse ESTs

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-resetc.riken.go.jp

Thermotabilization and thermoactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length cDNA  
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

Location/Qualifiers

1. .272

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="2010002H22"

/clone\_lib="Mus musculus small intestine C57BL/6J adult"

/sex="male"

FEATURES

source

Location/Qualifiers  
1. .230  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="1810061019"  
/clone\_lib="Mus musculus pancreas C57BL/6J adult"  
/sex="male"  
62 a 58 c 40 g 70 t

/tissue\_type="small intestine"  
/dev\_stage="adult"

77 a 60 c 45 g 90 t

BASE COUNT

ORIGIN

Query Match 1.9%; Score 20; DB 15; Length 272;

Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 214 taaagtgaagtctggggagcc 233

Db 131 TAAAGTGAGTCTGGGGAGCC 112

|||||

RESULT 6

AQ300268/c

LOCUS

DEFINITION

AQ300268

ACCESSION

AQ300268

VERSION

AQ300268.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 295)

AUTHORS

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and  
Hood,L.

TITLE

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE

99380589

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3170 row: P column: 3

Class: BAC ends

High quality sequence stop: 295.

Location/Qualifiers

1. .295

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=3170 Col=3 Row=P"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in  
E-Coli DH10B"

51 a 115 c 41 g 75 t 13 others

BASE COUNT

ORIGIN

Query Match 1.9%; Score 20; DB 91; Length 295;

Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 778 tctgatgctcccaataaagt 797

Db 21 TCTGATGCTCCCAATAAAGT 2

|||||

```

VERSION      BE003749.1  GI:8263982
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 372)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
              Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV3-BN0096-200
              400-162-d03&t3=2000-04-20&t4=1)
              Seq primer: puc 18 forward
              High quality sequence start: 87
              High quality sequence stop: 198.
FEATURES     Location/Qualifiers
              1..372
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="BN0096"
                /dev_stage="Adult"
                /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                Site_2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
                low stringency conditions."
BASE COUNT   115 a 72 c 61 g 124 t
ORIGIN
Query Match      1.9%; Score 20; DB 32; Length 372;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 cttaaaagcccaattggtg 272
      |||||||
Db 160 CTTAAAGCCCAATTCGATG 179

RESULT 8
LOCUS      W42753 421 bp mRNA EST 10-OCT-1996
DEFINITION zc24g06.s1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA
            clone IMAGE:323290 3' similar to PIR:A53664 A53664 epidermal
            surface antigen - human ;, mRNA sequence.
ACCESSION  W42753
VERSION     W42753.1  GI:1327213
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 421)
REFERENCE   1 (bases 1 to 421)
AUTHORS     Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman

```

```

.M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
R., Williamson,A., Wohldmann,P. and Wilson,R.
The Wasio-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
Insert length: 611 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 300.
FEATURES     Location/Qualifiers
              1..421
                /organism="Homo sapiens"
                /db_xref="GDB:1254802"
                /db_xref="taxon:9606"
                /clone="IMAGE:323290"
                /clone_lib="Soares_senescent_fibroblasts_NbHSF"
                /tissue_type="senescent_fibroblast"
                /lab_host="DH10B (ampicillin resistant)"
                /note="vector: pT73D (Pharmacia) with a modified
                polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI
                ; 1st strand cDNA was primed with a Not I - oligo(dT)
                primer [5'
                TGTACCAATCTGAAGTGGGAGCGCGCATTTTTTTTTTTT 3'],
                double-stranded cDNA was size selected, ligated to Eco RI
                adapters (Pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of a modified pT73 vector
                (Pharmacia). Library went through one round of
                normalization to a Cot = 5. Library constructed by Bento
                Soares and M.Fatima Bonaldo."
BASE COUNT   120 a 95 c 123 g 80 t
ORIGIN
Query Match      1.9%; Score 20; DB 40; Length 421;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 cactcaggtactgagcaga 202
      |||||||
Db 380 CACTCAGGTAAGTACTGAGCAGA 361

RESULT 9
LOCUS      AQ315018 594 bp DNA GSS 04-MAY-1999
DEFINITION RPC111-9408.TV RPC1-11 Homo sapiens genomic clone RPC1-11-9408, DNA
            sequence.
ACCESSION  AQ315018
VERSION     AQ315018.1  GI:4046481
KEYWORDS    GSS.
SOURCE      human.
ORGANISM     Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 594)
REFERENCE   1 (bases 1 to 594)
AUTHORS     Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
            Use of human BAC End Sequences for Sequence-Ready Map Building
            Unpublished (1998)
            Other_GSSs: RPC111-9408.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850

```

Tel: 301 838 0200  
Fax: 301 838 0208

Email: hbeetigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: T7  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..594

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="RPCI-11"

/sex="Male"

/cell\_type="Lymphocytes"

/note="Vector: pBACE3 6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"

138 a 125 c 105 g 226 t

## BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 20; DB 91; Length 594;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

767 agtagaactgtctgtgct 786

|||||

Db 304 AGTAGAACTGTCTGTGCT 323

## RESULT 10

AZ015292

LOCUS

DEFINITION

AZ015292

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other\_GSSs: RPCI-23-264H17.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (<http://inforesgen.com>). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 264 row: H column: 17  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..621

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="RPCI-23-264H17"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3 6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

175 a 137 c 85 g 220 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 20; DB 113; Length 621;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 gagaaactgtctgtgct 750

|||||

Db 214 GAGAACTTGTCTGTGAT 233

## RESULT 11

AW360531

LOCUS

DEFINITION

AW360531

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2221

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 6H7053 row: F column: 01.

Location/Qualifiers

1..625

/organism="Zea mays"

/cultivar="Illinois High Oil"

/db\_xref="taxon:4577"

/clone\_lib="687 - Early embryo from Delaware"

/tissue\_type="embryo"

/dev\_stage="14, 21, 28, and 35 days after pollination"

/lab\_host="E. coli SOUR"

/note="Organ: embryo; Vector: pBluescript SK; Site\_1: XhoI; Site\_2: EcoRI; Library was prepared by Stratagene using the Uni-ZAP XR system (Stratagene BN937328-12). Clones were picked by a O-bot after blue/white selection (ampicillin resistance - use 100 micrograms/microliter). Developed from a pool of equal amounts of RNA from pollinating embryos sampled at 14, 21, 28 and 35 days after pollination of the Illinois High Oil Maize strain Cycle 90. This closed strain has been selected for high oil concentration for 90 generations and originates from the 1890s era open pollinated variety Burr's White"

128 a 171 c 177 g 149 t

BASE COUNT

ORIGIN



Mon Nov 6 10:14:03 2000

Query Match 1.8%; Score 20; DB 21; Length 625;  
 Best Local Similarity 100.0%; Pred. No. 8.2;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 tctgggagcctcgaggatt 242  
 |||||  
 Db 318 TCTGGGAGCCTCGAGGATT 337

RESULT 12  
 A0106028 259 bp DNA GSS 28-AUG-1998  
 LOCUS HS\_3054\_AL\_A02\_MR CIT Approved Human Genomic Sperm Library D Homo  
 DEFINITION sapiens genomic clone Plate-3054 Col-3 Row-A, DNA sequence.  
 ACCESSION A0106028  
 VERSION A0106028.1 GI:3481384  
 KEYWORDS GSS.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 259)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu

Sequence Tagged Connector  
 Plate: 3054 row: A column: 3  
 Class: BAC ends  
 High quality sequence stop: 259.  
 Location/Qualifiers  
 1..259  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Plate-3054 Col-3 Row-A"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in  
 E-Coli DH10B"

BASE COUNT 117 a 46 c 32 g 63 t 1 others  
 ORIGIN

Query Match 1.8%; Score 19; DB 89; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 921 gctcttaactgagattata 939  
 |||||  
 Db 71 GCTCTTAAGTGGAGATTATA 89

RESULT 13  
 BE142685/c 263 bp mRNA EST 21-JUN-2000  
 LOCUS BE142685  
 DEFINITION ILO-HT0156-051099-120-f12 HT0156 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE142685  
 VERSION BE142685.1 GI:8605406  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
AUTHORS

1 (bases 1 to 263)  
 Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.

## TITLE

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=42-IL0-HT0156-051  
 099-120-f12at3-1999-10-05&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 7  
 High quality sequence stop: 263.

FEATURES  
source

1..263  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HT0156"  
 /dev\_stage="Adult"  
 /note="Organ: head neck; Vector: puc18; Site:1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 85 a 41 c 71 g 66 t

BASE COUNT  
ORIGIN

Query Match 1.8%; Score 19; DB 33; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 646 ttaattatctgggata 664  
 |||||  
 Db 170 TTAATTATCTGGGATA 152

## RESULT 14

D18871 306 bp mRNA EST 12-DEC-1995  
 LOCUS MUSGS01053 Mouse 3'-directed Mus musculus domesticus cDNA clone  
 DEFINITION mc0861 3', mRNA sequence.  
 ACCESSION D18871  
 VERSION D18871.1 GI:1089504  
 KEYWORDS EST.  
 SOURCE western European house mouse.  
 ORGANISM Mus musculus domesticus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 306)  
 Kawamoto,S., Okubo,K., Yoshii,J., Katsuki,M. and Matsubara,K.  
 Analysis of gene expression in mouse embryogenesis by 3'-directed  
 cDNA sequencing  
 Unpublished (1995)  
 Contact: Kawamoto,S., Okubo,K., Yoshii,J., Katsuki,M. and Matsubara  
 ,K.

Institute for Cellular and Molecular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.

## FEATURES

source

1. 306  
Location/Qualifiers  
/organism="Mus musculus domesticus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10092"  
/clone="mc0861"  
/clone\_lib="Mouse 3'-directed"  
/tissue\_type="decidual tissue (day 6.5-8.5 of gestation)"

BASE COUNT 77 a 56 c 56 g 112 t 5 others

ORIGIN

Query Match 1.8%; Score 19; DB 36; Length 306;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 695 cagcttgctctgtgattg 713

Db 277 CAGCTTGCTCTGTGATTG 295

## RESULT 15

AI466102/C

LOCUS

DEFINITION AI466102 410 bp mRNA EST 09-MAR-1999  
IMAGE:1246341 5', mRNA sequence.

ACCESSION

AI466102

VERSION

AI466102.1

KEYWORDS

EST.

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 410)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

MGI:660029

This read is a RESEQUENCE of a previously sequenced mouse clone

This read has been verified (found to hit its original self in the

correct orientation)

Seq primer: -40Rp from Gibco.

Location/Qualifiers

1. 410

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:1246341"

/clone\_lib="Soares\_mammary\_gland\_NbMMG"

/sex="male"

/tissue\_type="mammary gland"

/dev\_stage="4 weeks"

/lab\_host="DH10B"

/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia

) with a modified polylinker; Site\_1: Not I; Site\_2: Eco

RI; 1st strand cDNA was primed with a Not I - oligo(dt)

primer [5'

TGTACCAATCTGAAGTGGAGCGCGCGAATGGTTTTTTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

RNA provided by Dr. Minoru Ko, Wayne State Univ. Library

constructed and normalized by Bento Soares and M.Fatima

Bonald.

BASE COUNT 133 a 94 c 100 g 82 t 1 others

ORIGIN

## Query Match

Best Local Similarity 1.8%; Score 19; DB 10; Length 410;  
Matches 19; Conservative 100.0%; Pred. No. 29;

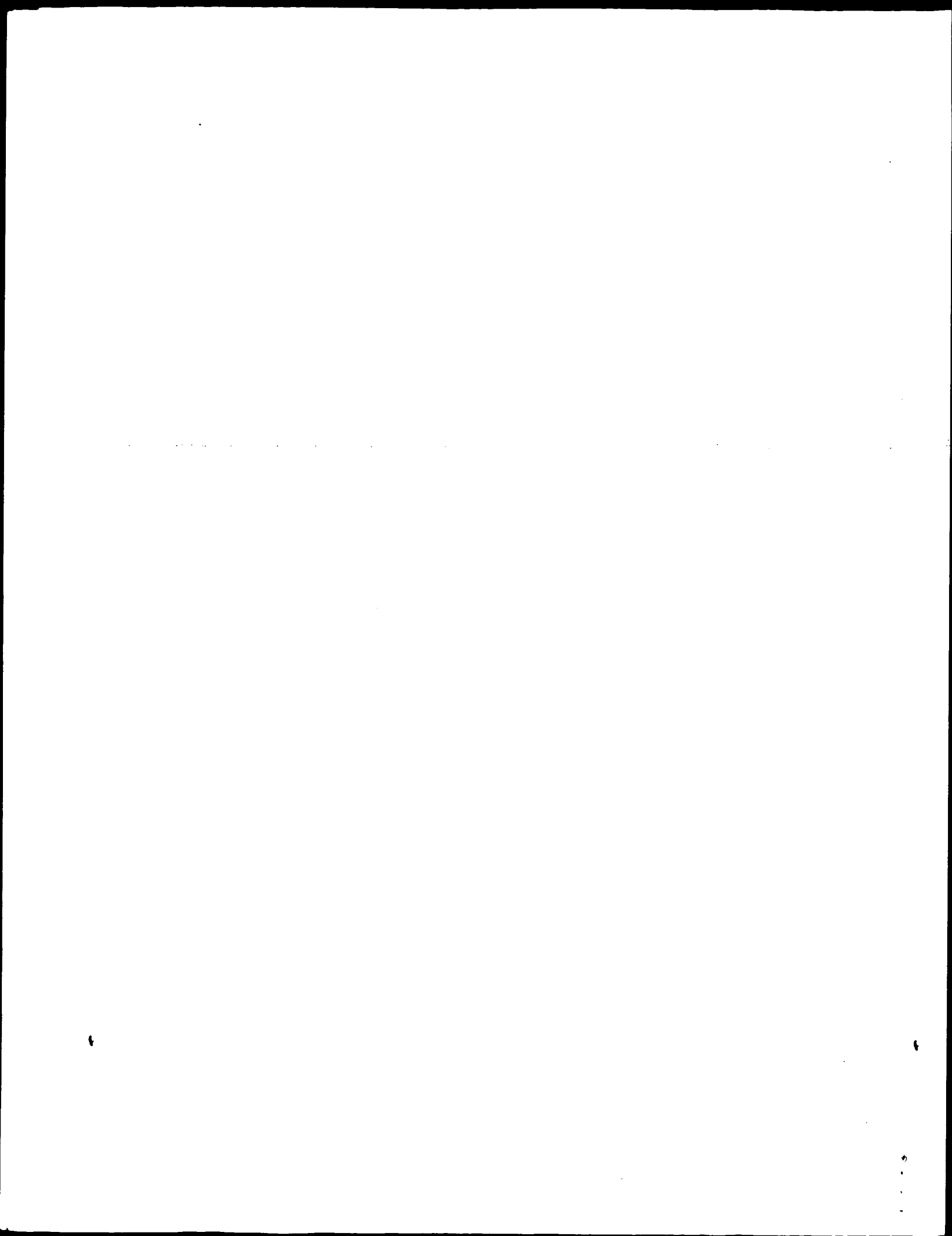
Mismatches 0; Indels 0; Gaps 0;

QY 104 cctgagtcctcttttgctt 122

Db 197 CCTGAGTCCCTTTTGCTT 179

Search completed: November 4, 2000, 01:08:36  
Job time: 17417 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2000, 17:59:19 ; Search time 7109.34 Seconds  
(without alignments)  
634.617 Million cell updates/sec

Title: US-09-101-423A-1  
Perfect score: 1033  
Sequence: 1 ctctctgggtctatgtc.....gtagacataactcccatggt 1033

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues 2067340  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl :

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_cm.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pl1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: em\_fun.\*
- 13: em\_hum1.\*
- 14: em\_hum2.\*
- 15: em\_in.\*
- 16: em\_om.\*
- 17: em\_ov.\*
- 18: em\_ov.\*
- 19: em\_pat.\*
- 20: em\_ph.\*
- 21: em\_pl.\*
- 22: em\_ro.\*
- 23: em\_sts.\*
- 24: em\_sy.\*
- 25: em\_un.\*
- 26: em\_vi.\*
- 27: gb\_htg1.\*
- 28: gb\_htg2.\*
- 29: gb\_in1.\*
- 30: gb\_in2.\*
- 31: em\_ba1.\*
- 32: em\_ba2.\*
- 33: em\_hum3.\*
- 34: em\_hum4.\*
- 35: gb\_pr4.\*
- 36: gb\_htg3.\*
- 37: gb\_htg4.\*
- 38: gb\_htg5.\*
- 39: gb\_htg6.\*
- 40: gb\_htg7.\*
- 41: em\_htg1.\*
- 42: em\_htg2.\*
- 43: em\_htg3.\*

- 44: em\_hum5.\*
- 45: gb\_pl3.\*
- 46: gb\_pr5.\*
- 47: gb\_htg8.\*
- 48: gb\_htg9.\*
- 49: gb\_htg10.\*
- 50: gb\_htg11.\*
- 51: gb\_htg12.\*
- 52: gb\_htg13.\*
- 53: gb\_htg14.\*
- 54: gb\_in3.\*
- 55: gb\_htg15.\*
- 56: gb\_htg16.\*
- 57: gb\_htg17.\*
- 58: em\_htg4.\*
- 59: em\_htg5.\*
- 60: em\_htg6.\*
- 61: em\_htg7.\*
- 62: em\_hum6.\*
- 63: gb\_htg18.\*
- 64: gb\_htg19.\*
- 65: gb\_ba3.\*
- 66: em\_htg8.\*
- 67: em\_htg9.\*
- 68: em\_htg10.\*
- 69: em\_htg11.\*
- 70: em\_htg12.\*
- 71: em\_htg13.\*
- 72: em\_htg14.\*
- 73: em\_htg15.\*
- 74: em\_htg16.\*
- 75: em\_htg17.\*
- 76: em\_htg18.\*
- 77: em\_htg19.\*
- 78: em\_htg20.\*
- 79: em\_htg21.\*
- 80: em\_htg22.\*
- 81: em\_htg23.\*
- 82: gb\_pr6.\*
- 83: gb\_pr7.\*
- 84: gb\_htg20.\*
- 85: gb\_htg21.\*
- 86: gb\_htg22.\*
- 87: gb\_htg23.\*
- 88: gb\_ro.\*
- 89: gb\_sts1.\*
- 90: gb\_sts2.\*
- 91: gb\_sy.\*
- 92: gb\_un.\*
- 93: gb\_vil.\*
- 94: gb\_vil2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1031.4	99.8	1033	5	A64058	A64058 Sequence 1
2	133.4	12.9	323501	52	AC027646	AC027646 Mus muscu
3	119.4	11.6	74332	57	AC073764	AC073764 Mus muscu
c 4	104.2	10.1	155063	50	AC025528	AC025528 Mus muscu
5	96.8	9.4	179252	88	AF131205	AF131205 Mus muscu
6	91.6	8.9	197819	57	AC073675	AC073675 Mus muscu
c 7	91.6	8.9	278652	57	AC073809	AC073809 Mus muscu
8	89	8.6	157647	56	AC069450	AC069450 Homo sapi
9	89	8.6	224142	50	AC025587	AC025587 Mus muscu
10	89	8.6	273800	88	AF100956	AF100956 Mus muscu
c 11	84.6	8.2	127934	57	AC073709	AC073709 Mus muscu
12	83.4	8.1	216295	56	AC069562	AC069562 Mus muscu



Mon Nov 6 10:14:04 2000

SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
 1 (bases 1 to 323501) Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,  
 Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,  
 Gordon, M., Goltz, J.S. and Kucherlapati, R.  
 High throughput Mouse Sequencing

TITLE  
 JOURNAL Direct Submission  
 AUTHORS 2 (bases 1 to 323501)  
 Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,  
 Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,  
 Gordon, M., Goltz, J.S. and Kucherlapati, R.

TITLE  
 JOURNAL Submitted (31-MAR-2000) Department of Molecular Genetics, Albert  
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
 Bronx, NY 10461, USA

COMMENT On Jun 9, 2000 this sequence version replaced gi:7543631.  
 -----Genome Center Center:  
 Albert Einstein College of Medicine Center  
 Code: ABCOM  
 Web site:  
<http://sequence.aecom.yu.edu/cgi-bin/ws.exe/mouseDB/mouseSQ/mouseseqtable.htm>  
 Contact: jhan@sequence.aecom.yu.edu  
 -----Summary Statistics  
 Center project name: ABE  
 Sequencing vector: pUC18; 108752  
 Chemistry: Dye-terminator Big Dye; 100%  
 Assembly program: Phrap version 0.990319  
 \*Consensus quality: 179084 at least Q20  
 \*Consensus quality: 171761 at least Q30  
 \*Consensus quality: 160720 at least Q40  
 \*\*Estimated insert size: agarose-FP - N/A  
 \*\*Estimated insert size: 319341 - sum-of-contigs  
 Quality coverage: agarose-FP - N/A  
 Quality coverage: 5.8x sum-of-contigs - N/A  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 209 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 656: contig of 656 bp in length  
 657 676: gap of unknown length  
 677 1357: contig of 681 bp in length  
 1358 1377: gap of unknown length  
 1378 2029: contig of 652 bp in length  
 2030 2049: gap of unknown length  
 2050 2867: contig of 818 bp in length  
 2868 2887: gap of unknown length  
 2888 3731: contig of 844 bp in length  
 3732 3751: gap of unknown length  
 3752 4431: contig of 680 bp in length  
 4432 4451: gap of unknown length  
 4452 5062: contig of 611 bp in length  
 5063 5082: gap of unknown length  
 5083 5983: contig of 901 bp in length  
 5984 6003: gap of unknown length  
 6004 6711: contig of 708 bp in length  
 6712 6731: gap of unknown length  
 6732 7410: contig of 679 bp in length  
 7411 7430: gap of unknown length  
 7431 8059: contig of 629 bp in length  
 8060 8079: gap of unknown length  
 8080 8829: contig of 750 bp in length  
 8830 8849: gap of unknown length  
 8850 9491: contig of 642 bp in length  
 9492 9511: gap of unknown length

9512 10417: contig of 906 bp in length  
 10418 10437: gap of unknown length  
 10438 11258: contig of 821 bp in length  
 11259 11278: gap of unknown length  
 11279 12103: contig of 824 bp in length  
 12103 12122: gap of unknown length  
 12123 12842: contig of 720 bp in length  
 12843 12862: gap of unknown length  
 12863 13717: contig of 855 bp in length  
 13718 13737: gap of unknown length  
 13738 14568: contig of 831 bp in length  
 14569 14588: gap of unknown length  
 14589 15351: contig of 763 bp in length  
 15352 15371: gap of unknown length  
 15372 16129: contig of 758 bp in length  
 16130 16449: gap of unknown length  
 16450 17080: contig of 931 bp in length  
 17081 17100: gap of unknown length  
 17101 17875: contig of 775 bp in length  
 17876 17895: gap of unknown length  
 17896 18473: contig of 578 bp in length  
 18474 18493: gap of unknown length  
 18494 19321: contig of 828 bp in length  
 19322 19341: gap of unknown length  
 19342 20020: contig of 679 bp in length  
 20021 20040: gap of unknown length  
 20041 20746: contig of 706 bp in length  
 20747 20766: gap of unknown length  
 20767 21586: contig of 820 bp in length  
 21587 22344: contig of 738 bp in length  
 22345 22364: gap of unknown length  
 22365 23093: contig of 729 bp in length  
 23094 23113: gap of unknown length  
 23114 23986: contig of 853 bp in length  
 23987 24844: contig of 858 bp in length  
 24845 24864: gap of unknown length  
 24865 25542: contig of 678 bp in length  
 25543 25562: gap of unknown length  
 25563 26443: contig of 881 bp in length  
 26444 26463: gap of unknown length  
 26464 27214: contig of 751 bp in length  
 27215 27234: gap of unknown length  
 27235 27899: contig of 665 bp in length  
 27900 27919: gap of unknown length  
 27920 28474: contig of 555 bp in length  
 28475 28494: gap of unknown length  
 28495 28908: contig of 414 bp in length  
 28909 28928: gap of unknown length  
 28929 29631: contig of 703 bp in length  
 29632 29651: gap of unknown length  
 29652 29956: contig of 305 bp in length  
 29957 29976: gap of unknown length  
 29977 30799: contig of 823 bp in length  
 30800 30819: gap of unknown length  
 30820 31516: contig of 697 bp in length  
 31517 31536: gap of unknown length  
 31537 32184: contig of 648 bp in length  
 32185 32204: gap of unknown length  
 32205 32838: contig of 634 bp in length  
 32839 32858: gap of unknown length  
 32859 33613: contig of 755 bp in length  
 33614 33633: gap of unknown length  
 33634 34338: contig of 705 bp in length  
 34339 34358: gap of unknown length  
 34359 35346: contig of 988 bp in length  
 35347 35366: gap of unknown length  
 35367 36189: contig of 823 bp in length  
 36190 36209: gap of unknown length  
 36210 36897: contig of 688 bp in length  
 36898 36917: gap of unknown length  
 36918 37581: contig of 664 bp in length

```
* 37582 37601: gap of unknown length
* 37602 38548: contig of 947 bp in length
* 38549 38568: gap of unknown length
* 38569 39143: contig of 575 bp in length
* 39144 39163: gap of unknown length
* 39164 39901: contig of 738 bp in length
* 39902 39921: gap of unknown length
* 39922 40725: contig of 804 bp in length
* 40726 40745: gap of unknown length
* 40746 41672: contig of 927 bp in length
* 41673 41692: gap of unknown length
* 41693 42385: contig of 673 bp in length
* 42386 42386: gap of unknown length
* 42387 42985: contig of 600 bp in length
* 42986 43005: gap of unknown length
* 43006 43488: contig of 483 bp in length
* 43489 43508: gap of unknown length
* 43509 44194: contig of 686 bp in length
* 44195 44214: gap of unknown length
* 44215 44939: contig of 725 bp in length
* 44940 44959: gap of unknown length
* 44960 45675: contig of 716 bp in length
* 45676 45695: gap of unknown length
* 45696 46228: contig of 533 bp in length
* 46229 46249: gap of unknown length
* 46250 46986: contig of 738 bp in length
* 46987 47006: gap of unknown length
* 47007 47659: contig of 653 bp in length
* 47660 47679: gap of unknown length
* 47680 48441: contig of 762 bp in length
* 48442 48461: gap of unknown length
* 48462 49334: contig of 873 bp in length
* 49335 49354: gap of unknown length
* 49355 50093: contig of 739 bp in length
* 50094 50113: gap of unknown length
* 50114 50778: contig of 665 bp in length
* 50779 50798: gap of unknown length
* 50799 51495: contig of 697 bp in length
* 51496 51515: gap of unknown length
* 51516 52215: contig of 700 bp in length
* 52216 52235: gap of unknown length
* 52236 53026: contig of 791 bp in length
* 53027 53047: gap of unknown length
* 53048 53625: contig of 579 bp in length
* 53626 53645: gap of unknown length
* 53646 54456: contig of 821 bp in length
* 54457 54467: gap of unknown length
* 54468 55212: contig of 726 bp in length

Query Match 12.9%; Score 133.4; DB 52; Length 323501;
Best Local Similarity 74.6%; Pred. No. 6.3e-30;
Matches 211; Conservative 0; Mismatches 61; Indels 11; Gaps 3;

QY 534 ttctcagatcacagaattattttgagctgttttttcatgtcccttgaggatg----- 587
Db 154378 TTCGGAACACAGAGATTTCTTGGATATATTTTTCATGCTCCCTCAGGTATAGCAA 154437

QY 588 ---ggagtgatttacttcagggttaacacagcatgtctactgttggtattcaagtaaatg 644
Db 154438 GTAGGAGTGAAGTTTACTTCATGTTACTCATTTATGTTGCTGCTACTTATCAATTAATG 154497

QY 645 cttaaatatccctggatataccctctgtggga-agcagggtttttgtatcatatcagcttgc 703
Db 154498 CCAGACAATTCCTTATCATGCCCTATGCAAGAATGATATTTTGCCTACATGGGCTTGC 154557

QY 704 ctgtgattg-atactgcttgactcaagaaacttgctcatgtgattcttcttaacog 762
Db 154558 ATTGTCAATGTTATACAGCTTGAACTCATGAGACTTTTACTCATGTCACCATCTTAAACC 154617

QY 763 atggagtagaanaactgtctgatgctctcaataaagttggctctt 805
Db 154618 TCAGAGAATAAATATCTGGTCTCTAAATAAAGTTGGCTATT 154660
```

AC073764 74332 bp DNA HTG 29-JUN-2000  
Mus musculus clone RP23-34809, WORKING DRAFT SEQUENCE, 28 unordered pieces.  
AC073764 1 GI:8810381  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 74332)  
DOE Joint Genome Institute.  
Sequencing of Mouse  
Unpublished  
2 (bases 1 to 74332)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 1870849  
Center clone name: RPCI-23\_34809  
-----  
Summary Statistics  
Consensus quality: 55837 bases at least Q40  
Consensus quality: 65003 bases at least Q30  
Consensus quality: 66938 bases at least Q20  
Estimated insert size: 214000; agarose-fp estimation  
Quality coverage: 4.75 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 14.2 in Q20 bases; agarose-fp estimation  
\* NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
1 1292: contig of 1292 bp in length  
\* 1293 1392: gap of unknown length  
\* 1393 2550: contig of 1158 bp in length  
\* 2551 2650: gap of unknown length  
\* 2651 4444: contig of 1794 bp in length  
\* 4445 4544: gap of unknown length  
\* 4545 6021: contig of 1477 bp in length  
\* 6022 6121: gap of unknown length  
\* 6122 7157: contig of 1036 bp in length  
\* 7158 7257: gap of unknown length  
\* 7258 8757: contig of 1500 bp in length  
\* 8758 8857: gap of unknown length  
\* 8858 10287: contig of 1430 bp in length  
\* 10288 10388: gap of unknown length  
\* 10389 11607: contig of 1220 bp in length  
\* 11608 11708: gap of unknown length  
\* 11709 12898: contig of 1191 bp in length  
\* 12899 12999: gap of unknown length  
\* 12999 14867: contig of 1869 bp in length  
\* 14868 14967: gap of unknown length  
\* 14968 16465: contig of 1498 bp in length  
\* 16466 16565: gap of unknown length  
\* 16566 18140: contig of 1575 bp in length  
\* 18141 18240: gap of unknown length  
\* 18241 19520: contig of 1280 bp in length  
\* 19521 19620: gap of unknown length



```
* 19621 20725: contig of 1105 bp in length
* 20726 20825: gap of unknown length
* 20826 22732: contig of 1907 bp in length
* 22733 22832: gap of unknown length
* 22833 25194: contig of 2362 bp in length
* 25195 25294: gap of unknown length
* 25295 27503: contig of 2209 bp in length
* 27504 27603: gap of unknown length
* 27604 30082: contig of 2479 bp in length
* 30083 30182: gap of unknown length
* 30183 31491: contig of 1309 bp in length
* 31492 31591: gap of unknown length
* 31592 34816: contig of 3225 bp in length
* 34817 34916: gap of unknown length
* 34917 38480: contig of 3564 bp in length
* 38481 38580: gap of unknown length
* 38581 44175: contig of 5595 bp in length
* 44176 44275: gap of unknown length
* 44276 49325: contig of 4950 bp in length
* 49326 49325: gap of unknown length
* 49326 53681: contig of 4356 bp in length
* 53682 53781: gap of unknown length
* 53782 58336: contig of 4535 bp in length
* 58337 58436: gap of unknown length
* 58437 63391: contig of 4855 bp in length
* 63392 63391: gap of unknown length
* 63392 69817: contig of 6426 bp in length
* 69818 69917: gap of unknown length
* 69918 74332: contig of 4415 bp in length.
FEATURES             Location/Qualifiers
     source           1..74332
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="RP23-34809"
BASE COUNT  19431 a 16668 c 16582 g 18905 t 2746 others
ORIGIN

Query Match      11.5%; Score 119.4; DB 57; Length 74332;
Best Local Similarity 65.6%; Pred. No. 1.2e-25;
Matches 238; Conservative 0; Mismatches 116; Indels 9; Gaps 4;

QY  509 agaaacagagctgtgacccctcattctcaagatcacagaattattcttgactgtgt 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  13920 ATAGACAGAGCTTATGACCTCAGTTCTTTAAACACAGGATTGTTGGNATGTGA 13979
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  569 ttcattgctctc-----tgttgatggagtgagtttacttcaggttaatacagcattgct-- 621
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  13980 TTCGTGCTCCTCCACGCCCTACCAATAGTGTGCTCTCTCAGATTACTCATTTATGCTGC 14039
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  622 tactgttggtattcaagtaaatccttaaatattcctcggatatacctctgtgggaag-cag 680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  14040 TGCTGTTAATGCAATAAAACGCTTTAAACCTCTCTTTATATGCTCTATGGAACACAT 14099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  681 gttttgatacatgcagctgtctgtgattgatactgcttgaactcaagagaacttgg 740
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  14100 GTTTTTCGCCACAGCTGCTGTCTGTGTGACCAAAACAGCAGCTTGAATTCGGAGAACTTTA 14159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  741 ctcattgattctttcttaaccgtaggagtagagaactgtctgatgctctcaataaaattgg 800
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  14160 CCCATGTGACCTTTCTTTAAACCCCTCAGGTA-TAGCTGTGCAGGAGCTCTGAATAAAGCTGG 14218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  801 ctcttgacagacggttagctgtctctgtttattctctccattcttcgcctccacggcc 860
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  14219 CTATTACATGAGACTTTTAGTCTGCCCTCATTAACAGGCTCCATTTGCCAGGTCCTCACTGTC 14278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  861 tct 863
    |||||
Db  14279 TCT 14281
    |||||
RESULT 4
```

```
* 77665 81704: contig of 4040 bp in length
* 81724: gap of unknown length
* 81725: contig of 5209 bp in length
* 86934: gap of unknown length
* 86934: gap of unknown length
* 90276: contig of 3323 bp in length
* 90296: gap of unknown length
* 90297: contig of 4664 bp in length
* 94960: gap of unknown length
* 94981: contig of 3854 bp in length
* 98834: gap of unknown length
* 98835: gap of unknown length
* 102475: contig of 3621 bp in length
* 102476: gap of unknown length
* 102496: contig of 4002 bp in length
* 106498: gap of unknown length
* 106517: gap of unknown length
* 110518: contig of 4066 bp in length
* 110584: gap of unknown length
* 110603: gap of unknown length
* 112934: contig of 2331 bp in length
* 112935: gap of unknown length
* 112954: gap of unknown length
* 115889: contig of 2935 bp in length
* 115909: gap of unknown length
* 115910: contig of 2954 bp in length
* 118864: gap of unknown length
* 118883: gap of unknown length
* 122889: contig of 4005 bp in length
* 122908: gap of unknown length
* 125440: contig of 2532 bp in length
* 125441: gap of unknown length
* 125461: contig of 3949 bp in length
* 129410: gap of unknown length
* 129429: gap of unknown length
* 132981: contig of 3552 bp in length
* 133001: gap of unknown length
* 133002: contig of 1376 bp in length
* 134378: gap of unknown length
* 134398: contig of 2278 bp in length
* 136675: gap of unknown length
* 136695: gap of unknown length
* 140904: contig of 4209 bp in length
* 140925: gap of unknown length
* 143584: contig of 2660 bp in length
* 143585: gap of unknown length
* 143605: contig of 2277 bp in length
* 145881: gap of unknown length
* 145901: gap of unknown length
* 145902: contig of 2976 bp in length
* 148877: gap of unknown length
* 148878: gap of unknown length
* 148898: contig of 3085 bp in length
* 151983: gap of unknown length
* 152003: contig of 3061 bp in length.
FEATURES
    source
    1..155063
    /organism="Mus musculus"
    /strain="C57BL6/J"
    /db_xref="taxon:10090"
    /chromosome="2"
    /clone="RP23-79p14"
    /sex="male"
BASE COUNT 46339 a 31972 c 32034 g 43891 t 827 others
ORIGIN
Query Match 10.1%; Score 104.2; DB 50; Length 155063;
Best Local Similarity 68.1%; Pred. No. 7e-21;
Matches 194; Conservative 0; Mismatches 78; Indels 13; Gaps 3;
Qy 534 ttctcaagatacagaattattcttggaactgtgttttcttcactgctccttggtgatg----- 587
Db 9164 TTCCTGGAACACACAGATTTGTTCTGGAATATGTTTTCATGCTCCTCCAGGTATAGCAA 9105
Qy 588 ---ggagtgaatttcacitcaggttaacacagcattgcttactgttggtattcaagtaaatg 644
Db 9104 GTAGAGGGAGTTTACTTCATGTTTACTCATATGATGGCTTGTGACTTATTCAATAAATG 9045
Qy 645 cttaaatattctggtatataacctgtggga-agcagggtttttgatacatgcagctgtgc 703
Db 9044 CCTAGACAAATCTTTACATGCCCTATGCAAGAATGATATTTTTCATACATGCGGGTGGG 8985
```

```
Qy 704 ctgtgatt---gatactgcttggaactcaagagaactttgtcatgtgattctttaac 760
Db 8984 AATTGCCATTGCAATACAGCTGGAACTAATGAAACACTTACTCATGTGACCATCTTAAC 8925
Qy 761 cgatggagtajaaactctgtatgctctcaataaaagtggctctt 805
Db 8924 CCTCAGAGAATAAATATATCTGGGTCTCTAAATAAAAGTTGGCTATT 8880
RESULT 5
AF131205 179252 bp DNA ROD 08-OCT-1999
LOCUS Mus musculus Serf1 protein (Serf1), survival of motor neuron
DEFINITION protein (Snm), neuronal apoptosis inhibitory protein-rs6
(Naip-rs3), and neuronal apoptosis inhibitory protein-rs3
(Naip-rs3) genes, complete cds.
ACCESSION AF131205
VERSION AF131205.1 GI:5932000
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 179252)
AUTHORS Endrizzi,M., Huang,S., Scharf,J.M., Kelter,A.R., Wirth,B.,
Kunkel,L.M., Miller,W. and Dietrich,W.F.
TITLE Comparative Sequence Analysis of the Mouse and Human Lgnl/SMA
Interval
JOURNAL Genomics 60 (2), 137-151 (1999)
REFERENCE 2 (bases 1 to 179252)
AUTHORS Dietrich,W.F., Endrizzi,M., Huang,S., Scharf,J. and Kunkel,L.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1999) Genetics, HHMI/Harvard Medical School, 200
Longwood Avenue, Boston, MA 02115, USA
FEATURES
    source
    1..179252
    /organism="Mus musculus"
    /strain="129/Sv"
    /db_xref="taxon:10090"
    /map="chromosome 13"
    /clone="BAC 149mi9"
    /rpt_family="B3A"
    /note="similar to Tfmr"
    /complement(825..972)
    /rpt_family="RSINE1"
    /complement(1176..1286)
    /rpt_family="B1-F"
    /complement(1488..1595)
    /rpt_family="RSINE1"
    /rpt_family="B1-MM"
    /rpt_family="B1-MM"
    /complement(3213..3314)
    /rpt_family="B3"
    /rpt_family="ID5"
    /rpt_family="B3"
    /complement(4044..4305)
    /rpt_family="HSMAR1"
    /rpt_family="HSMAR1"
    /complement(4937..5088)
    /rpt_family="HSMAR1"
    /rpt_family="B3"
    /complement(4421..4752)
    /rpt_family="HSMAR1"
    /complement(4825..4867)
    /rpt_family="HSMAR1"
    /complement(4937..5088)
    /rpt_family="HSMAR1"
    /rpt_family="B3"
    /complement(5102..5137)
    /rpt_family="B3"
```

```
repeat_region 5138..5480 /rpt_family="MTD"
repeat_region 5496..5655 /rpt_family="B3"
repeat_region complement(5660..5852) /rpt_family="HSMARI"
repeat_region 5854..6021 /rpt_family="LIME3"
repeat_region complement(6279..6380) /rpt_family="B2"
repeat_region 6759..6950
repeat_region /rpt_family="B3"
repeat_region 7295..7424 /rpt_family="B1-F"
repeat_region 7459..7671 /rpt_family="B3"
repeat_region 7686..7825 /rpt_family="B1-F"
repeat_region 8430..8553 /rpt_family="RSINEL"
repeat_region 8719..8795 /rpt_family="B1-F"
repeat_region 8834..8937 /rpt_family="PBID10"
repeat_region 9298..9508 /rpt_family="B3"
repeat_region 9517..9689 /rpt_family="B2"
repeat_region 9788..9843 /rpt_family="B2"
repeat_region 10004..10103 /rpt_family="RSINEL"
repeat_region complement(10605..11220) /rpt_family="RMER19"
repeat_region complement(11260..11310) /rpt_family="PBID10"
repeat_region complement(11345..11484) /rpt_family="B4A"
repeat_region complement(11519..11615) /rpt_family="PBID10"
repeat_region 11674..11750 /rpt_family="PBID7"
repeat_region complement(11947..12084) /rpt_family="B1-F"
repeat_region complement(12131..12211) /rpt_family="PB1"
repeat_region 12586..12770 /rpt_family="B3"
repeat_region 12778..12995 /rpt_family="B2"
repeat_region 13342..13471 /rpt_family="B1_MM"
repeat_region 15062..15543 /rpt_family="LX6"
repeat_region 15559..15750 /rpt_family="B2"
repeat_region complement(17531..17655) /rpt_family="B1-F"
repeat_region 17883..17959 /rpt_family="B1_MM"
repeat_region complement(18795..18951) /rpt_family="B1_MM"
repeat_region 19001..19197 /rpt_family="B2"
repeat_region complement(19397..19528) /rpt_family="PBID10"
repeat_region 19808..19972 /rpt_family="MTA"
repeat_region 19973..20114 /rpt_family="B1_MM"
repeat_region 20126..20345 /rpt_family="B1_MM"
repeat_region 20596..20695
```

```
repeat_region /rpt_family="B1-F"
repeat_region 20751..20871 /rpt_family="RSINEL"
repeat_region complement(21077..21419) /rpt_family="RMER15"
repeat_region 21427..21562 /rpt_family="B1-F"
repeat_region complement(21582..21670) /rpt_family="RMER15"
repeat_region complement(21831..21974) /rpt_family="B1_MM"
repeat_region 22038..22184 /rpt_family="B4A"
repeat_region 24337..24444 /rpt_family="PBID7"
repeat_region 24451..24604 /rpt_family="B3A"
repeat_region 24847..24967 /rpt_family="B1_MM"
repeat_region 25032..25232 /rpt_family="B3"
repeat_region 25253..25384 /rpt_family="B1-F"
repeat_region 25533..25698 /rpt_family="B2"
repeat_region 25701..25847 /rpt_family="B1_MM"
repeat_region 25947..26042 /rpt_family="B2"
repeat_region 26052..26249 /rpt_family="B2"
repeat_region 26254..26395 /rpt_family="L1MA5"
repeat_region 26421..26611 /rpt_family="B3"
repeat_region 26903..27038 /rpt_family="B4A"
repeat_region 26982..27038 /rpt_family="PBID7"
repeat_region 27134..27220 /rpt_family="PBID10"
repeat_region complement(27704..27806) /rpt_family="PBID9"
repeat_region 28122..28126 /rpt_family="BC1_MM"
repeat_region 28127..28186 /rpt_family="ID4"
repeat_region 28187..28205 /rpt_family="BC1_MM"
repeat_region complement(29256..29398) /rpt_family="B1_MM"
repeat_region complement(29399..29452) /rpt_family="B4"
repeat_region complement(29421..29504) /rpt_family="ID4"
repeat_region 29569..29672 /rpt_family="PBID10"
repeat_region complement(29741..29862) /rpt_family="PBID10"
repeat_region complement(29923..30025) /rpt_family="B3A"
repeat_region 30532..30550 /rpt_family="B4"
repeat_region 30551..30645 /rpt_family="PBID7"
```

```
Query Match 9.4%; Score 96.8; DB 88; Length 179252;
Best Local Similarity 73.5%; Pred. No. 1.4e-18;
Matches 150; Conservative 0; Mismatches 52; Indels 2; Gaps 2;
Qy 656 ctctgtgtgggaagcaggtttttgatacatgcagctgttcctgtgttg-atactgttga 724
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68086 CTCATAGAAAAACATGTTGCCACATGCAACTTGCTTCATGATGTATAGCCTGA 68145
```

```

QY 725 actcaagaacatttgcctatgtgatcttcttaaccgagtgagtagaagaactgtctgatg 784
||||| ||||||| ||||| ||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 68146 ACTCAGAGAAGACTTTCTCAGGTAGTTTCTTAACCTCAGAGATATAAAGCTGCTCATG 68205
QY 785 ctctcaataaagttggctcttcgcagagacacttgctgtctcttcttatctgtccattc 844
||||| ||||||| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 68206 CTCTGAATAAAGCTGACATATTGATGAGACTTTAGTCTGCTCATTTATTGGCTCTATT 68265
QY 845 ttcc-gctccacagcgcctctacag 867
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 68266 TCCCAAGTCTACTGCTCTAGAG 68289

RESULT 6
AC073675 197819 bp DNA HTG 18-JUL-2000
LOCUS Mus musculus clone RP23-100M13, WORKING DRAFT SEQUENCE, 11 ordered
DEFINITION pieces.
ACCESSION AC073675
VERSION AC073675.2 GI:9256753
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 197819)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 197819)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:8810292.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1775573
Center clone name: RPCI-23_100M13
-----
Summary Statistics
Consensus quality: 190968 bases at least Q40
Consensus quality: 195665 bases at least Q30
Consensus quality: 198434 bases at least Q20
Estimated insert size: 179000; agarose-fp estimation
Quality coverage: 10.63 in Q20 bases; sum-of-contigs estimation
Quality coverage: 9.64 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 14306: contig of 14306 bp in length
* 14307 14406: gap of unknown length
* 14407 49785: contig of 35379 bp in length
* 49786 49885: gap of unknown length
* 49886 72977: contig of 23092 bp in length
* 72978 73078: gap of unknown length
* 73078 78598: contig of 5521 bp in length
* 78598 104248: gap of unknown length
* 104248 104349: gap of unknown length
* 104349 105977: contig of 1629 bp in length

```

```

* 105978 106077: gap of unknown length
* 106078 107053: contig of 976 bp in length
* 107054 128517: gap of unknown length
* 107134 128517: contig of 21364 bp in length
* 128518 128617: gap of unknown length
* 128618 149845: contig of 21228 bp in length
* 149846 149845: gap of unknown length
* 149946 192788: contig of 42843 bp in length
* 192739 192888: gap of unknown length
* 192839 197819: contig of 4931 bp in length.
FEATURES
Location/Qualifiers
1..197819
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-100M13"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 50815 a 45613 c 47297 g 53094 t 1000 others
ORIGIN

Query Match 8.9%; Score 91.6; DB 57; Length 197819;
Best Local Similarity 66.2%; Pred No. 6.1e-17;
Matches 147; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 625 tgttgatttcaagtaaatgcttaattatctctggtatataacctctgtgggaagcaggttt 684
||||| ||||||| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 115981 TGCTGTGTTCAATTGAAAGTTTAAACTATGCTATATATATGCTTCCATTGGAAGCGGTG 116040
QY 685 ttgatacatgagctgtctctgtgtgattg-atactgtgtgaactcaagaagaacttggctc 743
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116041 TGTGACACTTGTGCTTGTGTTGTTGTCATTGTATACGGCTTGAGTTCTCGAGAGCTGTACTC 116100
QY 744 atgtgattcttcttaaccgagtgagtagaagaactgtctgtctctcaataaagttggctc 803
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116101 AGGTAGCACTTCTTAACTCCTCAGAGATATAAATTGACTCAGGCTCTGTATAACATGGCTA 116160
QY 804 ttgcacgagagctgtgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtct 845
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 116161 TTGTTATGAGACATTAGTCCAGCACTTATATTGGCTCCGCTCT 116202

RESULT 7
AC073809/c
LOCUS Mus musculus clone RP23-56M17, WORKING DRAFT SEQUENCE, 36 unordered
DEFINITION pieces.
ACCESSION AC073809
VERSION AC073809.1 GI:8810426
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 278652)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 278652)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1775573
Center clone name: RPCI-23_56M17
-----
Summary Statistics
Consensus quality: 252289 bases at least Q40

```

Consensus quality: 263895 bases at least Q30  
 Consensus quality: 266429 bases at least Q20  
 Estimated insert size: 215000; agarose-fp estimation  
 Quality coverage: 9.55 in Q20 bases; sum-of-contigs estimation  
 Quality coverage: 7.47 in Q20 bases; sum-of-contigs estimation  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 36 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1044: contig of 1044 bp in length  
 1045 1144: gap of unknown length  
 1145 2389: contig of 1245 bp in length  
 2390 2489: gap of unknown length  
 2490 3623: contig of 1134 bp in length  
 3624 3723: gap of unknown length  
 3724 4950: contig of 1227 bp in length  
 4951 5050: gap of unknown length  
 5051 6358: contig of 1308 bp in length  
 6359 6459: gap of unknown length  
 6460 7691: contig of 1232 bp in length  
 7692 9195: contig of 1405 bp in length  
 9196 9295: gap of unknown length  
 9296 10451: contig of 1155 bp in length  
 10452 11948: contig of 1398 bp in length  
 11949 12048: gap of unknown length  
 12049 13214: contig of 1166 bp in length  
 13215 13314: gap of unknown length  
 13315 14472: contig of 1158 bp in length  
 14473 14572: gap of unknown length  
 14573 15895: contig of 1323 bp in length  
 15896 15995: gap of unknown length  
 15996 17073: contig of 1078 bp in length  
 17074 17173: gap of unknown length  
 17174 18506: contig of 1333 bp in length  
 18507 18606: gap of unknown length  
 18607 20221: contig of 1615 bp in length  
 20222 20321: gap of unknown length  
 20322 21675: contig of 1354 bp in length  
 21676 21775: gap of unknown length  
 21776 23648: contig of 1873 bp in length  
 23649 23748: gap of unknown length  
 23749 25978: contig of 2230 bp in length  
 25979 26078: gap of unknown length  
 26079 28123: contig of 2045 bp in length  
 28124 28223: gap of unknown length  
 28224 30661: contig of 2438 bp in length  
 30662 30761: gap of unknown length  
 30762 33105: contig of 2344 bp in length  
 33106 33205: gap of unknown length  
 33206 37129: contig of 3924 bp in length  
 37130 37239: gap of unknown length  
 37240 40019: contig of 2790 bp in length  
 40020 40119: gap of unknown length  
 40120 44618: contig of 4499 bp in length  
 44619 44718: gap of unknown length  
 44719 50463: contig of 5745 bp in length  
 50464 50563: gap of unknown length  
 50564 58865: contig of 8302 bp in length  
 58866 58965: gap of unknown length  
 58966 73624: contig of 14729 bp in length  
 73625 7394: gap of unknown length  
 7395 86431: contig of 12637 bp in length  
 86432 86532: gap of unknown length  
 86533 97864: contig of 11333 bp in length  
 97865 119089: contig of 21125 bp in length

\* 119090 119189: gap of unknown length  
 \* 119190 137271: contig of 18082 bp in length  
 \* 137272 137371: gap of unknown length  
 \* 137372 134642: contig of 17271 bp in length  
 \* 134643 134742: gap of unknown length  
 \* 134743 177021: contig of 22279 bp in length  
 \* 177022 177121: gap of unknown length  
 \* 177122 200830: contig of 23709 bp in length  
 \* 200831 200930: gap of unknown length  
 \* 200931 238314: contig of 37384 bp in length  
 \* 238315 238414: gap of unknown length  
 \* 238415 278652: contig of 40238 bp in length.

FEATURES  
 source

1. .278652  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-56M17"  
 /clone\_lib="RPCI mouse BAC library 23"  
 BASE COUNT 68127 a 71133 c 70594 g 65253 t 3545 others  
 ORIGIN

Query Match 8.9%; Score 91.6; DB 57; Length 278652;  
 Best Local Similarity 66.2%; Pred. No. 6.4e-17;  
 Matches 147; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 625 tgggtgattcaagtaaatgcttaattatctggtatatactctgtgggaagcaggttt 684  
 DB 68717 TGCTGTGTTCAATGAAAGTTAAACTATGCTATATATGCTTCCATTTGGAACGGGGTG 68658  
 QY 685 ttgacacagcagcttgcctgtgattg-atactgttgtaactcaagagaactttgtctc 743  
 DB 68657 TGTGACACTTGTGCTTGTCTGTGTCATGTATACGGTCTGAGTTCCTGAGAGCTGTACTC 68598  
 QY 744 atgtgatctttcttaacagatggagtagaactgtctgagtctctcaataaagtgtctc 803  
 DB 68597 AGGTAGCATTTCTTAACCCCTCAGAGTATAAATGACTCAGGCTCTGTATACATTGGCTA 68538  
 QY 804 ttgacagagacgttagctgtcctgtttatctgtctccattct 845  
 DB 68537 TGTATGAGACATTAGTCCAGCACTTATATTTGGCTCGCTCT 68496

RESULT 8

AC069450 157647 bp DNA HTG 15-JUN-2000  
 LOCUS Homo sapiens clone GS3-445D10, WORKING DRAFT SEQUENCE, 15 unordered  
 DEFINITION pieces.

AC069450:  
 AC069450.4 GI:8493548

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 157647)

AUTHORS

Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,  
 Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,  
 Buha, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,  
 Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,  
 Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,  
 Gorrell, J. H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,  
 Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,  
 Kovar, C., Liu, J., Liu, W., Loulseghe, H., Lozano, R. J., Martin, R.,  
 Massey, E., McLeod, M. P., Mei, G., Moore, S., Morgan, M., Morris, S.,  
 Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,  
 Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,  
 Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,  
 Muzny, D. M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,  
 Worley, K. and Gibbs, R.

Direct Submission

JOURNAL

Unpublished





```

/clone_lib="Genome Systems"
<1..2000
/gene="DAXX"
join(<146..208,325..544,1825..2000)
/gene="DAXX"
/product="Fas-binding protein Daxx"
join(<146..208,325..544,1825..1884)
/gene="DAXX"
/notes="intron-exon boundaries defined in relation to cDNA
sequence in GenBank Accession Number AF006040"
/codon_start=2
/product="Fas-binding protein Daxx"
/protein_id="AAC69891.1"
/db_xref="GI:3811375"
/translation="PPSKRREKKKOLGSLGNSVIKEPMAQDQSGQNTSQVPMPS
PLASVASVADSSPRVDSPEHLYTSLCSFSLLLQTFQAQSLRQCIIYKTSVATQCD
PEIIIVLSDSD"
724..858
/rpt_family="MMB1"
complement(1257..1297)
/rpt_family="(CA)n"
1303..1421
/rpt_family="MMB1F"
complement(1426..1630)
/rpt_family="B3"
complement(1646..1792)
/rpt_family="MMB1F"
complement(1728..2788)
/rpt_family="GC-rich"
3188..3214
/rpt_family="MIR"
<3365..>5281
/gene="BING1"
/product="Bing1"
<3365..>5281
/gene="BING1"
/notes="Zn finger; mouse orthologue to BING 1, described in
GenBank Accession Number Z97183"
3365..5281
/gene="BING1"
/notes="Thisorf was predicted by Genscan and confirmed by
ESTs AA538332 and AA797063"
/codon_start=1
/product="Bing1"
/protein_id="AAC69892.1"
/db_xref="GI:3811376"
/translation="MEPSALSPSGATLPLPLSLAPPLPLPAAVHVHVSFPEVTSALL
ESLNQRLQGLQCDVIRYQGRFRAHRAVLAAASPYFDQVLLKMTSISLPSVMDP
GAFETVLASVYGRSLMAAADIVNLTGVSVLQMMHIVDKCTELLREGRSAAITTVT
TAAASVSPCASVPSGNGTVPATVGSVRSHTSSRASENOSPSSNYFSPRESTDF
SSTQDAFVASAAGSNRRDGGVFPVAPVYVAGTSGKLLLEADLQDGDGGRGAV
AFAGLRNRCNPASAVPQKHVVYKQARNCPAPASLVHQDPLEDEEEDLDLTCE
DDEDEMGGSVPAGEPEATLSIDVRTLPEADKGEQVNFCESSNDGPGVEGGG
ACAGLDPCGPTPSSVYALNHPRLPLLDVPGNQLVFPSSSQAPGPGPNTAEHGA
VILGTSVAGLIPSGSGAPAGTGNKIFLCKGKAFSHKSMRDRHVMHNLNR
PDCPCVNFKFMKHHLLTHMTHTGLKPYECVCAKFKPMWRDSFMRHRCRRHRM
GVGVSQPGPGPGPGTSPGALQPKRESTTGGGSGDEANSATPPSHRRVMSPPSVH
KVMDFSGGGGAH"
join(6121..6206,6353..6532,6731..6991,11888..12286,
12486..12827,13108..13184,14618..14754)
/gene="tapasin"
/product="tapasin"
<6121..>14754
/gene="tapasin"
join(6170..6206,6353..6532,6731..6991,11888..12286,
12486..12827,13108..13184,14618..14629)
/gene="tapasin"
/notes="intron-exon boundaries defined in relation to cDNA
in GenBank Accession Numbers AF043943 and EST AA562898;
EST W75528 contains an alternatively spliced form of this
transcript"
/codon_start=1
/product="tapasin"

```

---

```

/protein_id="AAC69893.1"
/db_xref="GI:3811377"
/translation="MKPELLLVAVALGLATVSVVSAGPEAIECFWEDAGGGGLSKK
PATLLRHGRGPPRPDLPKLYFKVDDPAGMLLAARFYPAGASAPHCMSRFIPF
PASAKWNSLSPQNCPRDALDGLWLVSVSTLFSLSLLRQPEPLREPVVITMATV
VLTVLTHNPAPRVOLGKDALDLRFVAPSALEGSPLDAGPPPEGLERROHKGKH
LLLAATPGLAGMPPAOEKATAFAAADDDEPGMTGNTFLPAVKPSOEGVYLATV
LLPLYQGVSLVLTTHKAPRVSLTPAPVYVNAAGAPPELLCLASLFFPAEGLEVKWE
LRGPGGSSRKVEGTWLSTIRHSDGVSOSGHLQPPVPTAKQGHVHYVCRVYHSSL
PASGRSADVTLEVAAYWTIPEVSEKAKATAASLTIPRNSKKSQ"
complement(8176..8383)
/rpt_family="B3"
8545..8670
/rpt_family="B3"
complement(9036..9243)
/rpt_family="MMB2"
complement(9853..9997)
/rpt_family="MMB1"
10003..10014
/gene="tapasin"
/notes="number of Ts unclear"
complement(10003..10198)
/rpt_family="MMB2"
10369..10460
/rpt_family="(GA)n"
10454..10474
/gene="tapasin"
/notes="low quality data"
complement(10836..10943)
/rpt_family="PB1D7"
10973..11020
/rpt_family="MUSID5"
11215..11328
/rpt_family="MMB1F"
11329..11427
/rpt_family="(GA)n"
complement(11428..11459)
/rpt_family="(CA)n"
11467..11637
/rpt_family="B3"
complement(13691..13830)
/rpt_family="MMB1"
complement(13859..13993)
/rpt_family="MMSQR2"
complement(14790..14932)
/rpt_family="MMB1F"
complement(14947..15087)
/rpt_family="MMB1"
complement(15009..15110)
/rpt_family="MTE2"
15133..15243
/rpt_family="MMB1"
complement(15387..15490)
/rpt_family="URR1"
join(16444..16522,16723..16919,18204..18287,18518..18696,
18799..18849,18932..19229,19487..19738,19829..19932,
20019..20103,20190..20258,20375..20449,21453..21485,
21612..21733,21838..21933,22023..22134,22218..22511,
23441..23455,23543..23754)
/gene="RLF"
/product="RafGDS-like factor"
16444..24873
/gene="RLF"
/notes="described as a Ras-associating protein in GenBank
Accession Number U54639"
join(16764..16919,18204..18287,18518..18696,18799..18849,
18932..19229,19487..19738,19829..19932,20019..20103,
20190..20258,20375..20449,21453..21485,21612..21733,
21838..21933,22023..22134,22218..22511,23341..23455,
23543..23754)
/gene="RLF"
/notes="intron-exon boundaries defined in relation to cDNA
found in GenBank Accession Number U54639"

```



```
/codon_start=1
/product="RalGDS-like factor"
/protein_id="AAC69894.1"
/db_xref="GI:3811378"
/translation="MLPRPLRLLLDTTPGGVVLVSFRSDPEEGDGGRAVGGQGE
EDEDEEEASVWDEEDGATFTVTSOYRDLPLDLPLPPRSRRRLRAGTLEALVR
HLIDARTAGADMFTPALATRAFTSTPALFYADLEALSHPPGELERTTGVAI
SVLSTWLAGHDFGSEVKGQDLRFLESTLLTYAARGVGVGSADLRNLRLRVDPRI
APDLKPLALPGSDPADTDLVFLADHLAEQLTLLDAELFLNLPSCGLGLWGHDRD
```

Query Match 8.6%; Score 89; DB 88; Length 273800;  
Best Local Similarity 60.8%; Pred. No. 4.1e-16;  
Matches 243; Conservative 0; Mismatches 110; Indels 48; Gaps 4;

QY 459 gaacatggtcacaagatcacagtcagtgagagaaatcccaatcgctcagagagaaacagag 518  
DB 97177 GGACATGGTCAGAGATGGAGTCTTTGAGGAGACTTCGAGTGCTCATGAGAAACTGT 97236

QY 519 agtctgtgacctcc--attctccaagatacacagaattattcttggactgtgtttctatgct 576  
DB 97237 AAGAAAACAAGTCTGTCTTCTCAAAACGGTGGAGTGTCTTGGATGTGTTTGAAGCT 97296

QY 577 cctt-----gtgatggagtgagtttactcaagtgtaatacgaattgcttactgt 627  
DB 97297 CATCCGGGCATAGTGGATAGGAGTACGTTACTTCAGATTATTCATTACAGCTGCTAT 97356

QY 628 tggattacaagtaaatcctaaattatcctggatatacctctgtggagacaggttttg 687  
DB 97357 TGATATTGTTTATACGCCCTGTGG-----GGAAACACGCTGCTG 97397

QY 688 atacatgagctgtctgtgtgactgatactgtgaactcaagagaaactttgctcatgt 747  
DB 97398 CTGATGTGGCCCTCTCTGTGCTGGAC-----ACTTACCACTGT 97439

QY 748 gatctttcttaacogatyggagtgaaactgtctgatgctctcaataagtggtcttgc 807  
DB 97440 GACTCTGCTTAACCTCAGAGTAAATGCTGATGCTGTAATAAGTGTGCTGTGC 97499

QY 808 acgagagctagctgtctgttttatctgtctcattcttcc 848  
DB 97500 AGGAGACTTAGTCCACCTCATTTTAAAGCTCTATTCTCC 97540

RESULT 11  
AC073709/c  
LOCUS AC073709 127934 bp DNA HTG 29-JUN-2000  
DEFINITION Mus musculus clone RP23-188122, WORKING DRAFT SEQUENCE, 38  
unordered pieces.  
ACCESSION AC073709  
VERSION AC073709.1 GI:8810326  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
UNPUBLISHED  
REFERENCE 2 (bases 1 to 127934)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 1809278  
Center clone name: RPCI-23\_188122  
-----

Summary Statistics  
Consensus quality: 104805 bases at least Q40  
Consensus quality: 114900 bases at least Q30  
Consensus quality: 117118 bases at least Q20  
Estimated insert size: 188000; agarose-fp estimation  
Estimated insert size: 124234; sum-of-contigs estimation  
Quality coverage: 7.0 in Q20 bases; agarose-fp estimation  
Quality coverage: 10.59 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 38 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1065: contig of 1065 bp in length  
1066: gap of unknown length  
1166: contig of 1351 bp in length  
2516: gap of unknown length  
2517: contig of 1154 bp in length  
3770: gap of unknown length  
3771: gap of unknown length  
3871: contig of 1031 bp in length  
4901: gap of unknown length  
5001: gap of 1506 bp in length  
5002: gap of unknown length  
5007: gap of unknown length  
6507: gap of 1453 bp in length  
6508: gap of unknown length  
8060: contig of 1453 bp in length  
8160: gap of unknown length  
9303: contig of 1145 bp in length  
9405: gap of unknown length  
10554: contig of 1149 bp in length  
10555: gap of unknown length  
11870: contig of 1216 bp in length  
11871: gap of unknown length  
11971: contig of 1048 bp in length  
13019: gap of unknown length  
13119: contig of 1004 bp in length  
14123: gap of unknown length  
14223: contig of 2216 bp in length  
14233: gap of unknown length  
14338: contig of 1580 bp in length  
16339: gap of unknown length  
18118: contig of 1580 bp in length  
18119: gap of unknown length  
18219: gap of 1226 bp in length  
19444: gap of unknown length  
19545: gap of unknown length  
21698: contig of 2153 bp in length  
21797: gap of unknown length  
21798: contig of 3333 bp in length  
25131: gap of unknown length  
25230: contig of 1675 bp in length  
25231: gap of unknown length  
26905: contig of 2092 bp in length  
26906: gap of unknown length  
27006: contig of 1520 bp in length  
29098: gap of unknown length  
29198: contig of 2577 bp in length  
30717: gap of unknown length  
30817: gap of 2577 bp in length  
33394: gap of unknown length  
33494: gap of unknown length  
33935: contig of 2336 bp in length  
35831: gap of unknown length  
35931: contig of 2738 bp in length  
38668: gap of unknown length  
38768: gap of unknown length  
38769: contig of 4275 bp in length  
43044: gap of unknown length  
43143: gap of unknown length  
43885: contig of 2742 bp in length  
45886: gap of unknown length  
45986: contig of 4624 bp in length  
50609: gap of unknown length  
50709: gap of unknown length  
55164: contig of 4455 bp in length  
55264: gap of unknown length  
55265: contig of 3952 bp in length  
59317: gap of unknown length  
65163: contig of 5847 bp in length  
65263: gap of unknown length  
73723: contig of 8460 bp in length

```

* 73724 73823: gap of unknown length
* 73824 78180: contig of 4357 bp in length
* 78181 78280: gap of unknown length
* 78281 84718: contig of 6438 bp in length
* 84719 89722: contig of 4904 bp in length
* 89723 96159: contig of unknown length
* 96160 96259: gap of unknown length
* 96260 103124: contig of 6865 bp in length
* 103125 103225: gap of unknown length
* 103226 107620: contig of 4396 bp in length
* 107621 107720: gap of unknown length
* 107721 114582: contig of 6862 bp in length
* 114583 114682: gap of unknown length
* 114683 120690: contig of 6008 bp in length
* 120691 120790: gap of unknown length
* 120791 127934: contig of 7144 bp in length.
FEATURES
    source
        1..127934
            Location/Qualifiers
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="RP23-188122"
                /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 35009 a 27401 c 27967 g 33852 t 3705 others
ORIGIN

Query Match      8.2%; Score 84.6; DB 57; Length 127934;
Best Local Similarity 52.4%; Pred. No. 8.7e-15;
Matches 251; Conservative 0; Mismatches 204; Indels 24; Gaps 2;

QY 434 cagctattagggaacacagctgtgtgaacatgttcccaagaatacacagtcagtgaggagaa 493
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51562 CGGTATTGTTGGAACTAAGAAATGGAGTCAGTTAGACTTCCTGAGGCTGTGTGAGAAAA 51503

QY 494 tcccaatgcgcagagagaacagagctgtgtgacctccattcttcaagatacagaatca 553
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51502 CTCCAAGAAAACAGACAGAAATAGTATCTGACCCAGTTTCCTCAAACTGCGAATTCG 51443

QY 554 ttcttgactgtgttttcactgtctctctctctctctctctctctctctctctctctctct 609
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51442 ATATTGGAGTGTGTTTGTGCCCCCTCCAGGTATACAGAGTGAGTATTATTCAGATTA 51383

QY 610 atcagcttctctctctctctctctctctctctctctctctctctctctctctctctct 669
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51382 TCAATTACAACTGCTACTGTATTGTTTATGTTGCTCAGAAAA----- 51336

QY 670 gtgggaagcaggtttttgatcacatgcagctgtgtctctgtgtgattgatactcttgaaactca 729
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51337 -----AAAACCATTTTGGCCATGAGTGGCTGTGCCCAAAATACGGCTTGTCTTGAT 51283

QY 730 agagaactttgctatgtatgtctctctctctctctctctctctctctctctctctctct 789
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51282 CGTACAGTTTGCATATGTGACTCTCTTGACCCCTCAGAGTAAAACTGTCGATGCTCTG 51223

QY 790 aataaatttgctcttcacagacagctgtctgtctctctctctctctctctctctctctct 849
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51222 ACTTAAGTGGCTATTGCATGAGACTTAGTCTGCTCACTATTGCTCCATCCCTCCCA 51163

QY 850 ctcccagggcctctacagcactaaacccacacacacacacacacacacacacacacacacac 908
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51162 GCTGAAGGAGCCCATATCTCTCTACACAGCGCATTAACAAAAAATATATCTGTATA 51104

RESULT 12
AC069562 216295 bp DNA HTG 15-JUL-2000
LOCUS Mus musculus chromosome 13 clone rp23-317el3 strain Black6, WORKING
DEFINITION DRAFT SEQUENCE, 18 unordered pieces.
ACCESSION AC069562
VERSION AC069562.10 GI:9211299
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

```

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Jul 15, 2000 this sequence version replaced gi:8714552.
* NOTE: this is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1977: contig of 1977 bp in length
* 1978 2077: gap of unknown length
* 2078 4286: contig of 2209 bp in length
* 4287 4386: gap of unknown length
* 4387 7233: contig of 2847 bp in length
* 7234 7333: gap of unknown length
* 7334 9851: contig of 2518 bp in length
* 9852 13794: contig of 3843 bp in length
* 13795 13894: gap of unknown length
* 13895 19424: contig of 5530 bp in length
* 19425 19525: gap of unknown length
* 19526 24730: contig of 5205 bp in length
* 24731 32885: contig of 8056 bp in length
* 32886 32985: gap of unknown length
* 32986 39818: contig of 6833 bp in length
* 39819 39918: gap of unknown length
* 39919 51241: contig of 11323 bp in length
* 51242 51341: gap of unknown length
* 51342 67797: contig of 16455 bp in length
* 67798 67896: gap of unknown length
* 67897 80195: contig of 12299 bp in length
* 80196 80295: gap of unknown length
* 80296 94088: contig of 13792 bp in length
* 94089 94187: gap of unknown length
* 94188 111951: contig of 17764 bp in length
* 111952 112051: gap of unknown length
* 112052 134021: contig of 21970 bp in length
* 134022 134121: gap of unknown length
* 134122 155343: contig of 21222 bp in length
* 155344 155443: gap of unknown length
* 155444 176590: contig of 21147 bp in length
* 176591 176690: gap of unknown length
* 176691 216295: contig of 39605 bp in length.
FEATURES
    source
        1..216295
            Location/Qualifiers
                /organism="Mus musculus"
                /strain="Black6"
                /db_xref="taxon:10090"
                /chromosome="13"
                /clone="rp23-317el3"
                /clone_lib="RPCI-23"
BASE COUNT 59779 a 47960 c 48467 g 58383 t 1706 others
ORIGIN

Query Match      8.1%; Score 83.4; DB 56; Length 216295;
Best Local Similarity 66.5%; Pred. No. 2.2e-14;
Matches 236; Conservative 0; Mismatches 91; Indels 28; Gaps 7;

```

QY 534 ttctcaagatacaagaattattcttgagctgttttctatgctccttggtgagtgagtg 593  
 Db 18329 TTTCTTCAAAACGAGTAGACTTATCTCGAATGTTGTTTGTGCTCTCCGGGTGTAGCAG 18388  
 QY 594 ag-----ttacttcagggttaatacagcattgcttactgtgtgtattcaagtaa 641  
 Db 18389 AGAGGAACGAGCGCTTCTCTCAGATTATCTGTTATTGCTTCTGCTGCTACTCGATTGC 18448  
 QY 642 atgcttaattatcctcctggtatcctcttggtggaagcaggtttttgata-----catgca 696  
 Db 18449 ATG-----TATCTCTTTATATACCTCTATGGAACACCATTTTGTCTACACGATGTA 18501  
 QY 697 gcttgccttgatgga-tactgcttgaaactcaagagaaacttgcctgctgctttc 755  
 Db 18502 CTTTGCCTTGATGCTTACGCTTGAACTCTGATTAACCTTACCCAGGTGACCTTTC 18561  
 QY 756 ttaaccgat-ggagtagaactgtctgctctcaataaaagtgtgcttgcacagac 814  
 Db 18562 TTTTACCCCTCAGAGGATAAAGTCTGCTGCTGCTGAATAAAGTGCACGCTGTGTGAGAC 18621  
 QY 815 -gttagctgctgttattctgctcattcttcc-gctccacagcgcctctacag 867  
 Db 18622 TTTTGTCCGAGTAACCTTACCCAGCTCCTCCAGGTCCACGCGCTCTAGAG 18676  
  
 RESULT 13  
 AC016087/c  
 LOCUS Homo sapiens clone RP11-24K13, WORKING DRAFT SEQUENCE, 38 unordered  
 DEFINITION pieces.  
 ACCESSION AC016087  
 VERSION AC016087.5 GI:6957544  
 KEYWORDS HTG; HGVS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 178016)  
 Birten.B., Linton.L., Nusbaum.C., Lander.E., Allen.N., Anderson.M.,  
 Baldwin.J., Barna.N., Beckerly.R., Boguslavskiy.L., Boukhalter.B.,  
 Brown.A., Castle.A., Colangelo.M., Collins.S., Collymore.A.,  
 Cooke.P., Dearellano.K., Dewar.K., Domino.M., Donelan.L., Doyle.M.,  
 Ferreira.P., Fitzhugh.W., Forrest.C., Funke.R., Gage.D.,  
 Galagan.J., Gardyna.S., Grant.G., Hagos.B., Heaford.A., Horton.L.,  
 Howland.J.C., Johnson.R., Jones.C., Kann.L., Karatas.A., Klein.J.,  
 Lehoczy.J., Lieu.C., Locke.K., Macdonald.P., Marquis.N.,  
 McEwan.P., McGurk.A., McKernan.K., McLaughlin.J., Meldrim.J.,  
 Morrow.J., Naylor.J., Norman.C.H., O'Connor.T., O'Donnell.P.,  
 Peterson.K., Pollara.V., Riley.R., Roy.A., Santos.R., Severy.P.,  
 Stange-Thomann.N., Stojanovic.N., Subramanian.A., Talamas.J.,  
 Testaye.S., Tirrell.A., Vassiliev.H., Vo.A., Wheeler.J., Wu.X.,  
 Wyman.D., Ye.W.J., Zimmer.A. and Zody.M.  
 Direct Submission  
 Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Feb 11, 2000 this sequence version replaced gi:6957531.  
 All repeats were identified using RepeatMasker:  
 Smit, A.P.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L4484  
 Center clone name: 24\_K13  
 ----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 154757 bases at least Q40  
 Consensus quality: 168173 bases at least Q30  
 Consensus quality: 173336 bases at least Q20  
 Insert size: 155000; agarose-fp  
 Quality coverage: 3.5 in Q20 bases; agarose-fp  
 Quality coverage: 3.1 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 38 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1024: contig of 1024 bp in length  
 \* gap of unknown length  
 \* 1025: contig of 2065 bp in length  
 \* gap of unknown length  
 \* 3091: contig of 1676 bp in length  
 \* gap of unknown length  
 \* 4767: contig of 1359 bp in length  
 \* gap of unknown length  
 \* 6126: contig of 1512 bp in length  
 \* gap of unknown length  
 \* 7637: contig of 1988 bp in length  
 \* gap of unknown length  
 \* 9626: contig of 481 bp in length  
 \* gap of unknown length  
 \* 10107: contig of 1603 bp in length  
 \* gap of unknown length  
 \* 11710: contig of 3208 bp in length  
 \* gap of unknown length  
 \* 14918: contig of 792 bp in length  
 \* gap of unknown length  
 \* 15710: contig of 2530 bp in length  
 \* gap of unknown length  
 \* 18240: contig of 2674 bp in length  
 \* gap of unknown length  
 \* 20914: contig of 4324 bp in length  
 \* gap of unknown length  
 \* 25238: contig of 2159 bp in length  
 \* gap of unknown length  
 \* 27397: contig of 2426 bp in length  
 \* gap of unknown length  
 \* 29823: contig of 3303 bp in length  
 \* gap of unknown length  
 \* 33126: contig of 4422 bp in length  
 \* gap of unknown length  
 \* 37548: contig of 2428 bp in length  
 \* gap of unknown length  
 \* 39976: contig of 3021 bp in length  
 \* gap of unknown length  
 \* 42997: contig of 3560 bp in length  
 \* gap of unknown length  
 \* 46557: contig of 6017 bp in length  
 \* gap of unknown length  
 \* 52574: contig of 5886 bp in length  
 \* gap of unknown length  
 \* 58460: contig of 5194 bp in length  
 \* gap of unknown length  
 \* 63654: contig of 5477 bp in length  
 \* gap of unknown length  
 \* 69131: contig of 5879 bp in length  
 \* gap of unknown length  
 \* 75010: contig of 5139 bp in length  
 \* gap of unknown length  
 \* 80149: contig of 6161 bp in length  
 \* gap of unknown length

```

* 86310 93469: contig of 7160 bp in length
*      gap of unknown length
* 93470 98982: contig of 5513 bp in length
*      gap of unknown length
* 98983 104921: contig of 5939 bp in length
*      gap of unknown length
* 104922 112526: contig of 7605 bp in length
*      gap of unknown length
* 112527 122037: contig of 9511 bp in length
*      gap of unknown length
* 122038 130295: contig of 8258 bp in length
*      gap of unknown length
* 130296 137837: contig of 7542 bp in length
*      gap of unknown length
* 137838 148168: contig of 10331 bp in length
*      gap of unknown length
* 148169 154467: contig of 6299 bp in length
*      gap of unknown length
* 154468 164476: contig of 10009 bp in length
*      gap of unknown length
* 164477 178016: contig of 13540 bp in length.
FEATURES
  source      Location/Qualifiers
              1..178016
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="RP11-24K13"
              /clone_lib="RP11-24K13"
              /clone_lib="RP11-24K13"
BASE COUNT  51839 a 36865 c 36521 g 52752 t 39 others
ORIGIN

Query Match      8.1%; Score 83.2; DB 39; Length 178016;
Best Local Similarity 61.6%; Pred. No. 2.5e-14;
Matches 196; Conservative 0; Mismatches 98; Indels 24; Gaps 3;

QY 534 tcttcagaatacagaattattcttggaactgtgttttcttcacgtcc-----ttgtgg 584
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165462 TTCTCAAAACATGGAATTCCTTTGGAGTGTGTTTGGAGTTCGCCAGGTTGGTAG 165403
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 585 atggaggtgagttactcattcaggtaatacagcattgcttactgtgttattcaagtaaatg 644
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165402 ATAGGAGTAGTTACTTACAGATTATCATATTGCTGCTGCTGCTTATTCATTACATG 165343
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 645 cttaaatattctggatataacctctgtggga-----agcagggttttgata 690
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165342 TTTAAGCTATCCATTATATGCTCTATGGAACACAGAGGCATAGTGTTTTGCCA 165283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 691 catcagcttgccttgat-tgatactgtgaactcaagaagaactttgctcatgta 749
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165282 TGTGCGGCTGTCTCTGTATCATATGACAGCTTGAATGCAATGCAATGCAATGCAATGCA 165223
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 750 tcttcttaacagatgagtagaactgtctgatgctctcaataaagttggcttgcac 809
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165222 TTATCTGTACCCAGCAGTAGCTGTGGAGTTGAATATGTCATGAGACTTTAGTCCACCTC 165163
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 810 gagacgttagtctgtcct 827
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165162 ATTTTGAAGCTTGTCTCT 165145
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
AC046150
LOCUS
DEFINITION  AC046150 176564 bp DNA HTG 30-MAY-2000
            Mus musculus chromosome 12 clone RP23-283B24, *** SEQUENCING IN
            PROGRESS ***, 64 unordered pieces.
ACCESSION  AC046150
VERSION    AC046150.3 GI:8103535
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 176564)

```

## AUTHORS

Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C., Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T., Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J., Bunay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z., Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R., Gorrell, J. H., Gunaratne, P., Haller, G., Hernandez, J., Hoques, M., Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S., Kovar, C., Liu, J., Liu, W., Loulsegod, H., Lozado, R. J., Martin, R., Massey, E., McLeod, M. P., Mei, G., Moore, S., Morgan, M., Morris, S., Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogihara, M., Parish, B., Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S., Williamson, A., Wrenford, G., Zhou, X., Bouck, J., Hodgson, A., Muzny, D. M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G., Worley, K., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 176564)  
Worley, K.C.  
Direct Submission  
Submitted (13-APR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 30, 2000 this sequence version replaced gi:7596296.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: MAEG  
Center clone name: RP23-283B24  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-terminator Big Dye; 90% of reads  
Chemistry: Dye-terminator Big Dye; 10% of reads  
Assembly program: Phrap; version 0.90611  
Consensus quality: 138594 bases at least Q40  
Consensus quality: 159021 bases at least Q30  
Consensus quality: 167637 bases at least Q20  
Estimated insert size: 170200; sum-of-contigs estimation  
Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

NOTE: This is a 'working draft' sequence. It currently consists of 64 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 7412: contig of 7412 bp in length  
7512: gap of unknown length  
7513 14047: contig of 6535 bp in length  
14048 14147: gap of unknown length  
14148 21165: contig of 7018 bp in length  
21166 21265: gap of unknown length  
21266 26979: contig of 5714 bp in length  
26980 27079: gap of unknown length  
27080 32628: contig of 5549 bp in length  
32629 32728: gap of unknown length  
32729 37514: contig of 4786 bp in length  
37515 37614: gap of unknown length  
37615 42853: contig of 5239 bp in length  
42854 42954: gap of unknown length  
42954 46705: contig of 3751 bp in length  
46705 46804: gap of unknown length  
46805 49877: contig of 3073 bp in length  
49878 49977: gap of unknown length  
49978 53810: contig of 3833 bp in length  
53811 53910: gap of unknown length  
53911 57731: contig of 3821 bp in length  
57732 57831: gap of unknown length



## RESULT 15

```

AC006949/c
LOCUS      AC006949 123423 bp DNA ROD 27-APR-1999
DEFINITION Mus musculus chromosome 6, clone 454_F_16, complete sequence.
ACCESSION  AC006949
VERSION     AC006949.8 GI:4689494
KEYWORDS   HTG.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 123423)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Mus musculus chromosome 6, clone 454_F_16
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 123423)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehocsky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyl,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
            Direct Submission
TITLE     Submitted (02-MAR-1999) Whitehead Institute/MIT Center for Genome
JOURNAL   Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE  3 (bases 1 to 123423)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehocsky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyl,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
            Direct Submission
TITLE     Submitted (27-APR-1999) Whitehead Institute/MIT Center for Genome
JOURNAL   Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Apr 27, 1999 this sequence version replaced gi:4580407.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES   source
            1..123423
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="454_F_16"
            /clone_lib="CITB mouse BAC library"
            /map="6"
            /chromosome="6"
            158..307
            /rpt_family="MIR"
            repeat_region
            1797..1830
            /rpt_family="(CA)n"
            repeat_region
            2158..2209
            /rpt_family="(CA)n"
            repeat_region
            2680..2964
            /rpt_family="Lx8"
            repeat_region
            complement(3129..3389)
            /rpt_family="B4A"
            repeat_region
            complement(4712..5064)
            /rpt_family="ORRIA3"
            repeat_region
            complement(5065..6840)
            /rpt_family="ORRIA-INT"
            repeat_region
            complement(6841..7192)
            /rpt_family="ORRIA3"
            repeat_region
            8610..8754
            /rpt_family="MIR"
            repeat_region
            9009..9180
            /rpt_family="(TTTC)n"
            repeat_region
            9181..9205
            /rpt_family="(TTCC)n"
            repeat_region
            9206..9387
            /rpt_family="(TTTC)n"
            repeat_region
            complement(10945..10995)
            /rpt_family="LMC1"
            repeat_region
            12327..12423
            /rpt_family="MTD"
            repeat_region
            12427..12473
            /rpt_family="CT-rich"
            repeat_region
            complement(12500..12658)
            /rpt_family="B1-F"
            repeat_region
            12674..12971
            /rpt_family="MTD"
            repeat_region
            complement(12972..13260)
            /rpt_family="B4A"
            repeat_region
            complement(13649..13824)
            /rpt_family="URR1B"
            repeat_region
            16479..16622
            /rpt_family="MIR"
            repeat_region
            17617..17822
            /rpt_family="Lx8"
            repeat_region
            17835..17879
            /rpt_family="(TAGA)n"
            repeat_region
            18180..18207
            /rpt_family="AT-rich"
            repeat_region
            18350..18719
            /rpt_family="Lx9"
            repeat_region
            18725..18845
            /rpt_family="URRIA"
            repeat_region
            18920..19431
            /rpt_family="Lx9"
            repeat_region
            19432..19466
            /rpt_family="(TCTG)n"
            repeat_region
            19467..19583
            /rpt_family="Lx9"
            repeat_region
            19599..19794
            /rpt_family="B2"
            repeat_region
            20220..20309
            /rpt_family="(TCTG)n"
            repeat_region
            20312..20362
            /rpt_family="(CA)n"
            repeat_region
            20578..20610
            /rpt_family="(TG)n"
            repeat_region
            20926..20953
            /rpt_family="(TTTG)n"
            repeat_region
            23095..25154
            /rpt_family="(TG)n"
            repeat_region
            25446..25662
            /rpt_family="MIR"
            repeat_region
            26529..26850
            /rpt_family="RLTRETN_XM"
            repeat_region
            26893..27048
            /rpt_family="ETnERV"
            repeat_region
            complement(27049..27098)
            /rpt_family="L2"
            repeat_region
            27099..28532
            /rpt_family="ETnERV"
            repeat_region
            28594..29058
            /rpt_family="(TTTC)n"
            repeat_region
            complement(29064..29208)

```



Mon Nov 6 10:14:04 2000

us-09-101-423a-1.rge

Page 20



us-09-101-423a-1.rng

Mon Nov 6 10:14:06 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2000, 18:32:10 ; Search time 254.92 Seconds  
(without alignments)  
1522.280 Million cell updates/sec

Title: US-09-101-423A-1  
Perfect score: 1033  
Sequence: 1 ctctcttggtgctctatgtc.....gtagacataactccatggt 1033

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues 960044

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1020.4	98.8	1034	18 T72785	Metastasis inducin
C 2	39.4	3.8	1053	19 V16956	Nucleotide sequenc
C 3	39.4	3.8	1339	19 V16954	Nucleotide sequenc
4	35.4	3.4	1526	17 T33520	Protein phosphatas
5	34.4	3.3	1823	19 T96323	S. pneumoniae deri
6	33.8	3.3	2492	16 Q79480	Thai catfish GHRH/
C 7	33	3.2	56516	20 Z00870	Pgi genomic coding
C 8	33	3.2	56520	20 Z01022	Wild type Pgi codi
C 9	32.8	3.2	5059	20 X84332	Stealth virus nucl
C 10	32.6	3.2	954	18 T90220	Magnetospirillum s
C 11	32.4	3.1	10461	20 X20553	Polynucleotide seq
12	32.2	3.1	133894	17 T13635	ACNPV genomic DNA

13	32	3.1	1642	21 Z35835	Partial nucleotide
14	32	3.1	1777	21 A12413	cDNA encoding a hu
15	32	3.1	1819	21 Z52195	Human transforming
16	32	3.1	2361	21 Z52201	Human transforming
17	32	3.1	2387	21 A27411	Corn histidyl-tRNA
C 18	31.6	3.1	402	20 A40586	Human secreted pro
19	31.4	3.0	741	20 Z16720	Human gene express
C 20	31.2	3.0	763	21 A02129	Human colon cancer
C 21	30.8	3.0	2962	20 X21408	targetting region
C 22	30.6	3.0	1980	19 V57559	DNA sequence of be
23	30.6	3.0	2321	21 Z51029	Pre-selection open
24	30.6	3.0	2337	21 Z51030	Pre-selection open
25	30.4	2.9	855	19 V04615	Flea serine protea
26	30.4	2.9	4698	13 Q32370	PIA genomic DNA.
27	30.4	2.9	4698	15 Q72475	Entire PIA genomic
28	30.4	2.9	4698	16 T01176	P815A antigen prec
29	30.4	2.9	4698	17 T14692	MAGE genomic DNA.
30	30.4	2.9	4698	20 X84111	PIA genomic sequen
31	30.2	2.9	300	21 A01055	Human colon cancer
C 32	30.2	2.9	1444	20 Z17016	Human gene express
33	30.2	2.9	1485	20 X15123	cDNA encoding huma
34	30.2	2.9	111309	20 X20250	Borrelia burgdorfe
C 35	30	2.9	438	20 X40635	Human secreted pro
C 36	30	2.9	462	20 X41224	Human prostate can
C 37	30	2.9	1024	21 Z97167	Human transport-as
C 38	30	2.9	1270	20 Z11732	Seq ID No: 24 of W
39	30	2.9	3324	20 X60557	S. aureus Sdr DNA
40	30	2.9	3354	20 X60556	Staphylococcus aur
41	30	2.9	3945	20 X77593	Borrelia burgdorfe
C 42	30	2.9	4005	18 V74942	H. pylori cell env
C 43	30	2.9	53585	20 X20251	H. pylori ORF hplp
C 44	29.8	2.9	984	18 V25041	
C 45	29.8	2.9	1029	18 V24826	

## ALIGNMENTS

RESULT 1  
T72785  
ID T72785 standard; DNA; 1034 BP.  
XX  
AC T72785;  
XX  
DT 22-SEP-1997 (first entry)  
XX Metastasis inducing DNA C2.  
DE Metastasis inducing DNA; Met-DNA; cancer; tumour; diagnosis;  
KW Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;  
KW osteopontin; ss.  
XX Homo sapiens.  
OS WO9725443-A1.  
PN 17-JUL-1997.  
PD 10-JAN-1997; 97WO-GB000074.  
PF 10-JAN-1996; 96GB-0000470.  
PR (UyLi-) UNIV LIVERPOOL.  
PA Barracough BR, Rudland PS;  
XX WPI; 1997-372878/34.  
XX New isolated metastasis-inducing DNA - used to develop products to  
PT identify and treat patients at risk from metastatic tumours  
PS Claim 8; page 25; 38pp; English.  
XX Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12  
CC

QY	1	cttccctgggtgctctatgctcttgcctctccctctctccagtcaccattaaagccataaccatt	60
Db	1	cttccctgggtgctctatgctcttgcctctccctctctccagtcaccattaaagccataaccatt	60
QY	61	cttgacagactctgggaacagtcctccctctgctctctctgttgccctcgtcagtcctcttttcg	120
Db	61	cttgacagactctgggaacagtcctccctctgctctctctgttgccctcgtcagtcctcttttcg	120
QY	121	ctggagacccttcacgtagctcccatctggatgacactagatagaaacgtggaagattgt	180
Db	121	ctggagacccttcacgtagctcccatctggatgacactagatagaaacgtggaagattgt	180
QY	181	caacatcagggttaactgacagcgtcagagatttaaagtgcctctgggagcctcgagga	240
Db	181	caacatcagggttaactgacagcgtcagagatttaaagtgcctctgggagcctcgagga	240
QY	241	ttagctctctcccttaaaagccaatggatgactaacccagactattgctacttttaggt	300
Db	241	ttagctctctcccttaaaagccaatggatgactaacccagactattgctacttttaggt	300
QY	301	gggaagtcaactagcatatctcgatgggtcacatctcgaaaggtttctagcagtggtggcc	360
Db	301	gggaagtcaactagcatatctcgatgggtcacatctcgaaaggtttctagcagtggtggcc	360
QY	361	tgtgtgagcagcattggtgtgtatcatggtgtgcagcatactcaggtcgtctgcaacct	420
Db	361	tgtgtgagcagcattggtgtgtatcatggtgtgcagcatactcaggtcgtctgcaacct	420
QY	421	cgaagcctcttcacgtatttaggggaacacactgggtgtt-gaacatggtccaaagatacag	479
Db	421	cgaagcctcttcacgtatttaggggaacacactgggtgtt-gaacatggtccaaagatacag	480
QY	480	tcatgtgaggagaatcccaatgcgtcaggagaaacagagatcgtgacctccattcttc	539
Db	481	tcatgtgaggagaatcccaatgcgtcaggagaaacagagatcgtgacctccattcttc	539
QY	540	aagatcacagaattattcttggactgtgtttcatgctctctctgttgatggaagagtgattta	599
Db	541	aagatcacagaattattcttggactgtgtttcatgctctctctgttgatggaagagtgattta	600
QY	600	cttcaggttaaacagcattgcttactgtgtgtattccaagttaaatgcttaaatctactcg	659
Db	601	cttcaggttaaacagcattgcttactgtgtgtattccaagttaaatgcttaaatctactcg	660
QY	660	atatacctctgtgggaacaggtttttgatacatgcagcttgcctctgtgattgatactg	719
Db	661	atatacctctgtgggaacaggtttttgatacatgcagcttgcctctgtgattgatactg	720
QY	720	cittgaactcaagagaactttgctcatgtgactttcttaaccgatgagtagaactgtc	779
Db	721	cittgaactcaagagaactttgctcatgtgactttcttaaccgatgagtagaactgtc	780
QY	780	tgatgctctcaataaagtggctcttgacagacgtttagctgtcctgttatctgtc	839
Db	780	tgatgctctcaataaagtggctcttgacagacgtttagctgtcctgttatctgtc	839











KW Treponema pallidum infection; syphilis; Borrelia infection; animal;  
 KW enzyme production; ds.  
 XX  
 OS Treponema pallidum.  
 XX  
 PN WO9859034-A2.  
 XX  
 PD 30-DEC-1998.  
 XX  
 PF 23-JUN-1998; 98WO-US13041.  
 XX  
 PR 24-JUN-1997; 97US-0050667.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Fraser CM;  
 XX  
 DR WPI; 1999-081273/07.  
 XX  
 PT New isolated Treponema pallidum nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of T. pallidum infections, particularly syphilis  
 XX  
 PS Claim 1; Page 491-497; 1150pp; English.  
 XX  
 CC X20500-21243 represent polynucleotide sequences from the genome of  
 CC Treponema pallidum. The sequences can be used for detection,  
 CC diagnosis, characterisation, prevention and therapy for T. pallidum  
 CC infections, particularly syphilis. They can also be used for detecting  
 CC diseases related to Borrelia infections in animals, and for the  
 CC production of biosynthetic products such as enzymes.  
 XX  
 SQ Sequence 10461 BP; 2325 A; 3411 C; 2493 G; 2213 T; 19 other;  
 3.1%; Score 32.4; DB 20; Length 10461;  
 Query Match Best Local Similarity 58.2%; Pred. No. 6.8;  
 Matches 57; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
 OY 670 gtgggaacagcttttggatcacatgcagctgtctctgtgattgatactgtgaactca 729  
 Db 1977 GTGGGGAAGAACAGCTTTTTCACCGCGCAGTGCCTTTGGGGTTGATAATCGGGCACTCT 1918  
 OY 730 agagaactttgctcatgtgactcttttcttaaccgatga 767  
 Db 1917 CGTGACCTCGGAATGATGCGCGTTCAGGTTGGA 1880  
 RESULT 12  
 T13635  
 ID T13635 standard; DNA; 133894 BP.  
 XX  
 AC T13635;  
 XX  
 DT 03-SEP-1996 (first entry)  
 XX  
 DE AcNPV genomic DNA clone 6.  
 XX  
 KW Autographa californica nuclear polyhedrosis virus clone 6;  
 KW disruption; non-essential gene; heterologous protein production;  
 KW expression vector; baculovirus; ss.  
 XX  
 OS Autographa californica nuclear polyhedrosis virus clone 6.  
 XX  
 PN WO9601320-A2.  
 XX  
 PD 18-JAN-1996.  
 XX  
 PF 30-JUN-1995; 95WO-IB00578.  
 XX  
 PR 04-JUL-1994; 94GB-0013420.  
 XX  
 PA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.

XX Ayres M, Bishop D, Possee R;  
 XX WPI; 1996-087670/09.  
 XX GENBANK; L22858.  
 XX  
 PT Autographa californica nuclear polyhedrosis virus complete genome  
 PT sequence - useful in the prodn. of vectors for enhanced  
 PT heterologous protein expression, such as interleukin(s),  
 PT interferon(s) and neurotoxin(s)  
 XX  
 PS Disclosure; Page 90-186; 122pp; English.  
 XX  
 CC The complete nucleotide sequence of the genome of clone 6 of the  
 CC baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV)  
 CC has been determined. The sequence is taken from the Genbank record  
 CC L22858. The patent specification claims a polynucleotide selected from  
 CC open reading frames (ORFs 13, 20, 22-26, 28-30, 32, 38, 41-45, 50-60,  
 CC 62-63, 66, 68-73, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,  
 CC 140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by  
 CC the patentees. See T13636-731. Expression vectors contg. the complete  
 CC genomic sequence of AcNPV, with the exception that at least one non-  
 CC essential ORF is disrupted or replaced are useful for the synthesis of  
 CC heterologous proteins.  
 XX  
 SQ Sequence 133894 BP; 39195 A; 27151 C; 27347 G; 40201 T; 0 other;  
 3.1%; Score 32.2; DB 17; Length 133894;  
 Query Match Best Local Similarity 66.7%; Pred. No. 32;  
 Matches 46; Conservative 0; Mismatches 23; Indels 0; Gaps 0;  
 OY 729 aagagaacttgcctcatgtgactcttttcttaaccgatgagtagaactgtctgtctct 788  
 Db 114227 acgagaacttctctcaaaagtatttttgcacgttgaaatgacacttgcattgtt 114286  
 OY 789 caataaagt 797  
 Db 114287 caataaact 114295  
 RESULT 13  
 Z36835  
 ID Z36835 standard; cDNA; 1642 BP.  
 XX  
 AC Z36835;  
 XX  
 DT 13-MAR-2000 (first entry)  
 XX  
 DE Partial nucleotide sequence of human interleukin-22.  
 XX  
 KW Human; interleukin-22; IL-22; IL-21; immune system disorder;  
 KW immune cell chemotaxis; haematopoietic cell disorder;  
 KW haemostatic activity; thrombolytic activity; autoimmune disorder; asthma;  
 KW respiratory problem; organ rejection; graft-versus-host disease; GVHD;  
 KW inflammation; hyperproliferative disorder; tissue regeneration;  
 KW embryonic stem cell differentiation; embryonic stem cell proliferation;  
 KW haematopoietic lineage; allergic asthma; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 3..485  
 FT /\*tag= a  
 FT /product= "partial IL-22"  
 XX  
 PN WO9961617-A1.  
 XX  
 PD 02-DEC-1999.  
 XX  
 PF 27-MAY-1999; 99WO-US11644.  
 XX  
 PR 29-MAY-1998; 98US-0087340.

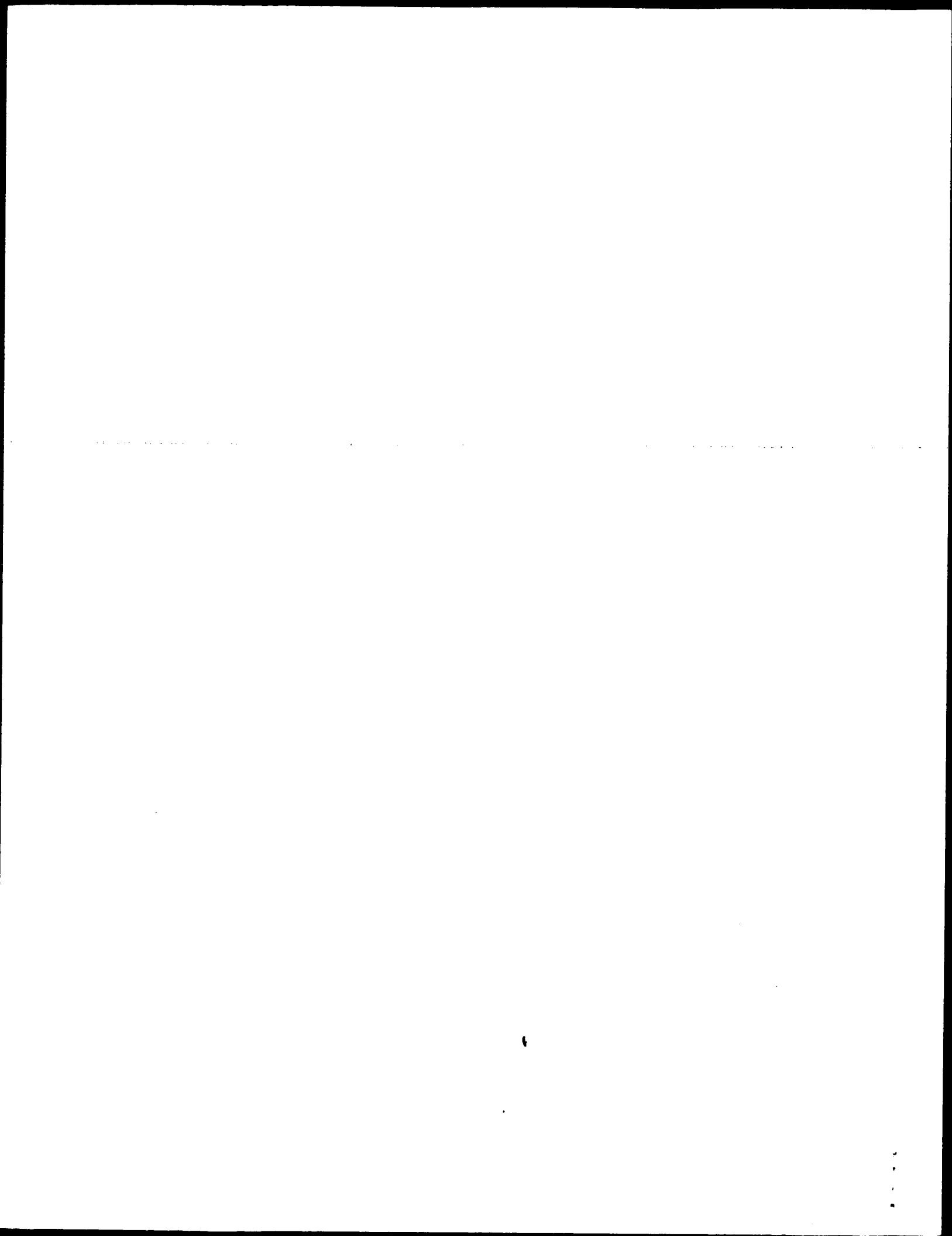


PR 10-SEP-1998; 98US-0099805.  
 PR 30-APR-1999; 99US-0131965.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 FI Ruben SM, Ebner R;  
 XX  
 XX WPI; 2000-072622/06.  
 DR P-PSDB; Y53891.  
 XX  
 XX Novel polynucleotides used to develop products for treating e.g. immune  
 PT disorders, blood disorders, autoimmune disorders, allergies,  
 PT inflammation, hyperproliferative disorders or infections -  
 XX  
 XX Claim 12; Fig 2A-B; 170pp; English.  
 XX  
 XX The present sequence encodes a partial human interleukin-22 (IL-22)  
 CC protein. The specification also describes IL-21 polynucleotides and  
 CC polypeptides. The IL-21 polynucleotide was isolated from a cDNA library  
 CC of epileptic frontal cortex. IL-21 and IL-22 may be useful in treating  
 CC deficiencies or disorders of the immune system, by activating or  
 CC inhibiting the proliferation, differentiation, or mobilization  
 CC (chemotaxis) of immune cells, treating or detecting deficiencies  
 CC of haematopoietic cells, to modulate haemostatic or thrombolytic  
 CC activity, in treating or detecting autoimmune disorders, treating  
 CC asthma (particularly allergic asthma) or other respiratory problems,  
 CC to treat and/or prevent organ rejection or graft-versus-host disease  
 CC (GVHD), to modulate inflammation, to treat or detect hyperproliferative  
 CC disorders, to treat or detect infectious agents, to differentiate,  
 CC proliferate and attract cells, leading to the regeneration of tissues,  
 CC IL-21 and IL-22 may also increase or decrease the differentiation or  
 CC proliferation of embryonic stem cells and haematopoietic lineage, may  
 CC be used to modulate mammalian characteristics.  
 XX  
 SQ Sequence 1642 BP; 377 A; 458 C; 439 G; 368 T; 0 other;  
 Query Match 3.1%; Score 32; DB 21; Length 1642;  
 Best Local Similarity 49.4%; Pred. No. 3.3;  
 Matches 83; Conservative 0; Mismatches 85; Indels 0; Gaps 0;  
 QY 49 agccataaccattctgacagactctggagacgtccctctctctctctgttggcctga 108  
 Db -  
 426 atcgacaacagggcgccaaagctctctgtggcccaacagcgcgcgcctga 485  
 QY gtcccttttgcctgagacccttcacgtacgtacgtccctccctctggtgacgtagaagac 168  
 Db -  
 486 ggccggtctctgcccgggaggtctcccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 545  
 QY gtgggaagtgtcacactcagtaactgagcagcagcagcagcagcagcagcagcagcagc 216  
 Db -  
 546 gctggagggtcgtgctggcgacctctgaagagagtgccacgagcagcagcagcagcagc 593  
 RESULT 14  
 ID AL2413  
 XX  
 AC AL2413;  
 XX  
 XX 25-JUL-2000 (first entry)  
 DT  
 DE CDNA encoding a human RNA-associated protein.  
 XX  
 KW Human; RNA-associated protein; cell proliferation; cancer; inflammation;  
 KW Immune response; reproductive disorder; actinic keratosis;  
 KW atherosclerosis; arteriosclerosis; bursitis; cirrhosis; hepatitis;  
 KW mixed connective tissue disease; myelofibrosis; primary thrombocythemia;  
 KW paroxysmal nocturnal hemoglobinuria; polycythemia vera; psoriasis;  
 KW trauma; ss.  
 XX  
 OS Homo sapiens.  
 XX

FH Key Location/Qualifiers  
 FT CDS 294...1520  
 FT /\*tag= a  
 XX /product= "RNA-associated protein"  
 PN WO200015799-A2.  
 XX  
 XX 23-MAR-2000.  
 PD  
 XX 17-SEP-1999; 99WO-US21688.  
 PF  
 XX 17-SEP-1998; 98US-0156039.  
 PR 22-SEP-1998; 98US-0158720.  
 PR 04-NOV-1998; 98US-0186815.  
 PR 08-APR-1999; 99US-0128660.  
 XX  
 XX (INCY-) INCYTE PHARM INC.  
 PA  
 XX Tang YT, Corley NC, Guegler KJ, Gorgone GA, Patterson C;  
 PI Hillman JL, Baughn MR, Lal P, Azlmzai Y, Yue H, Yang J;  
 DR WPI; 2000-271437/23.  
 DR P-PSDB; Y84441.  
 XX  
 XX New polypeptides and polynucleotides, useful for preventing and  
 PT treating a disorder associated with increased or decreased expression  
 PT of RNA associated proteins -  
 XX  
 XX Claim 9; Page 122; 131pp; English.  
 XX  
 XX The present sequence encodes a human RNA-associated protein. The  
 CC expression of RNA-associated proteins is closely associated with  
 CC reproductive tissues, nervous tissues, cell proliferation including  
 CC cancer, inflammation and immune responses, and so they may be used  
 CC for diagnosis, treatment or prevention of cell proliferative,  
 CC immune/inflammatory disorders, and reproductive disorders. Diseases  
 CC and disorders which may be treated include actinic keratosis,  
 CC atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis,  
 CC mixed connective tissue disease, myelofibrosis, paroxysmal nocturnal  
 CC hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia  
 CC and cancers, and trauma.  
 XX  
 SQ Sequence 1777 BP; 565 A; 329 C; 366 G; 517 T; 0 other;  
 Query Match 3.1%; Score 32; DB 21; Length 1777;  
 Best Local Similarity 46.8%; Pred. No. 3.5;  
 Matches 101; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
 QY 540 aagatacagattattcttggactgtgtttctcatgtctctctgtggatggagtgagttta 599  
 Db -  
 76 atgatagaagattttattcttgaaacgataataaagtgcagtcgtctgtctctcttcttca 135  
 QY 600 cttaaggttaatacagcattcttactgttggtattcaagtaaatgcttaaatctctgg 659  
 Db -  
 136 cccaagatcaagctgcagcattctccctccctccagagtcaggaagagtggttag 195  
 QY 660 atatacctctgtgggaagcaggtttttgatacatgcagcttgccttggtgactgactg 719  
 Db -  
 196 caaccaattgtcagagacgggtatctacttactgtgtgtgtgtgtgtgtgtgtgtgtgt 255  
 QY 720 ctggaactcaagagaactttgtctcatgtgtcttcttcttcttcttcttcttcttctt 755  
 Db -  
 256 gaagacaaagaaaaataagtaaccatgaagcagtc 291  
 RESULT 15  
 ID Z52195  
 XX Z52195 standard; cDNA; 1819 BP.  
 AC Z52195;  
 XX  
 XX 18-JUL-2000 (first entry)  
 DT







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2000, 18:03:16 ; Search time 141.08 Seconds  
(without alignments)  
1107.403 Million cell updates/sec

Title: US-09-101-423A-1

Perfect score: 1033

Sequence: 1 cttcttggtgtctatgtc.....gtagacataactcccatggt 1033

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/PCROS.COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36.2	3.5	7218	1	US-08-232-463-14
2	33.8	3.3	2502	2	US-08-062-472B-2
3	33	3.2	56516	4	US-08-996-306-1
4	32.6	3.2	954	4	US-08-973-275-2
5	32.4	3.1	7218	1	US-08-232-463-14
6	30.8	3.0	2962	3	US-08-189-256A-3
7	30.6	3.0	780	4	US-08-804-794-2
8	30.6	3.0	1980	4	US-08-804-794-1
9	30.4	2.9	855	5	US-08-906-769-130
10	30.4	2.9	855	5	US-08-906-616-130
11	30.4	2.9	4698	1	US-07-807-043B-5
12	30.4	2.9	4698	1	US-08-299-849B-5
13	30.4	2.9	4698	4	US-08-142-368A-5
14	30.4	2.9	4698	5	US-08-967-727-5
15	30.2	2.9	1485	4	US-08-861-512-2
16	29.8	2.9	9207	5	US-08-388-353-800
17	29.8	2.9	9207	5	US-08-488-551B-800
18	29.8	2.9	10607	2	US-08-078-090-3
19	29.8	2.9	35081	3	US-08-752-760A-1
20	29.4	2.8	2938	4	US-08-343-443B-3
21	29.2	2.8	2100	1	US-07-927-316A-1
22	29.2	2.8	3487	7	5164180-5
23	29.2	2.8	3487	7	5188960-1
24	29.2	2.8	3528	1	US-07-828-788A-11
25	29.2	2.8	3528	1	US-08-356-034-1
26	29.2	2.8	3528	5	US-08-933-891-1

Sequence 11, Appl  
Sequence 2, Appl  
Sequence 30, Appl  
Sequence 30, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 16, Appl  
Sequence 17, Appl  
Sequence 1, Appl  
Patent No. 5210189  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 20, Appl

Sequence 11, Appl  
Sequence 2, Appl  
Sequence 30, Appl  
Sequence 30, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 16, Appl  
Sequence 17, Appl  
Sequence 1, Appl  
Patent No. 5210189  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 20, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELE: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
; US-08-232-463-14

```

DB      247  TAGACACACA;ACACGCTCAGCAGCCGACCCCGAAGCCCGT 287
RESULT      3
US-08-996-306-1/c
: Sequence 1, Application US/08996306
: Patent No. 5945522
: GENERAL INFORMATION:
: APPLICANT: Cohan, Daniel
: APPLICANT: Chumakov, Ilya
: APPLICANT: Blumenfeld, Marta
: APPLICANT: Bougueleret, Lydie
: TITLE OF INVENTION: Prostate cancer gene
: NUMBER OF SEQUENCES: 68
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 501 West Broadway
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92101-3505
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy Disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: Win95
: SOFTWARE: Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/996,306
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelaelen, Ned A.
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: GENSET.018A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 56516 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: DOUBLE
: TOPOLOGY: LINEAR
: MOLECULE TYPE: GENOMIC DNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: Promoter
: LOCATION: 1629..1870
: IDENTIFICATION METHOD: Proscan
: FEATURE:
: NAME/KEY: Potential ATG
: LOCATION: 1998..2000
: FEATURE:
: NAME/KEY: Exon 1
: LOCATION: 2001..2216
: FEATURE:
: NAME/KEY: ATG
: LOCATION: 2031..2033
: FEATURE:
: NAME/KEY: TTA phos
: LOCATION: 11694..14332
: FEATURE:
: NAME/KEY: SEQ ID42
: LOCATION: 11930..11947
: FEATURE:
: NAME/KEY: SEQ ID24
: LOCATION: 12057..12103
: FEATURE:
: NAME/KEY: SEQ ID51
: LOCATION: compl(12339..12358)
: FEATURE:
: NAME/KEY: SEQ ID64

```

MOR NOV 6 10:14:07 2000

```
LOCATION: 13547..13564
FEATURE:
NAME/KEY: SEQ ID58
LOCATION: 13557..13703
FEATURE:
NAME/KEY: SEQ ID67
LOCATION: compl(13962..13981)
FEATURE:
NAME/KEY: Exon 2
LOCATION: 18196..18265
FEATURE:
NAME/KEY: Exon 3
LOCATION: 23717..23832
FEATURE:
NAME/KEY: Exon 4
LOCATION: 25571..25660
FEATURE:
NAME/KEY: SEQ ID43
LOCATION: 34216..34234
FEATURE:
NAME/KEY: SEQ ID25
LOCATION: 34469..34515
FEATURE:
NAME/KEY: SEQ ID52
LOCATION: compl(34625..34645)
FEATURE:
NAME/KEY: Exon 5
LOCATION: 34669..34759
FEATURE:
NAME/KEY: Exon 6
LOCATION: 40688..40846
FEATURE:
NAME/KEY: Exon 7
LOCATION: 48070..48193
FEATURE:
NAME/KEY: Exon 8
LOCATION: 50182..54523
FEATURE:
NAME/KEY: SEQ ID65
LOCATION: 51149..51168
FEATURE:
NAME/KEY: SEQ ID59
LOCATION: 51448..51494
FEATURE:
NAME/KEY: SEQ ID68
LOCATION: compl(51482..51499)
FEATURE:
NAME/KEY: SEQ ID44
LOCATION: 51596..51613
FEATURE:
NAME/KEY: SEQ ID26
LOCATION: 51612..51658
FEATURE:
NAME/KEY: SEQ ID53
LOCATION: compl(51996..52015)
FEATURE:
NAME/KEY: polyAd signal
LOCATION: 54445..54450
US-08-996-306-1
```

```
Query Match 3.2%; Score 33; DB 4; Length 56516;
Best Local Similarity 60.7%; Pred. No. 4.1;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 895 tcttccactgcaacacatccacagaggtctttaaactgagattataaaactgttactagatg 954
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38172 TCTCCAAAGAAATCCAGTCAAGACTCCACATTAAGATTATTGAATCAAAAGGGATA 38113

Qy 955 atgggttgaatcgctcccaagaacataa 983
      ||||| ||||| ||||| ||||| |||||
Db 38112 TTGGGTGCACTCTTTCCAGAGAAATATAA 38084
```

```
RESULT 4
US-08-973-275-2/c
; Sequence 2, Application US/08973275B
; Patent No. 5958706
; GENERAL INFORMATION:
; APPLICANT: MATSUNAGA, Tadashi
; APPLICANT: KAMIYA, Shinji
; APPLICANT: NAMEA, Kenyo
; TITLE OF INVENTION: FINE MAGNETIC PARTICLES CONTAINING USEFUL PROTEINS
; TITLE OF INVENTION: BOUND THERETO, PROCESS FOR PRODUCING THE SAME, AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: MATSUNAGA
; CURRENT APPLICATION NUMBER: US/08/973,275B
; CURRENT FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: PCT/JP97/01043
; EARLIER FILING DATE: 1997-03-27
; EARLIER APPLICATION NUMBER: JP 8-97536
; EARLIER FILING DATE: 1996-03-28
; EARLIER APPLICATION NUMBER: JP 8-146833
; EARLIER FILING DATE: 1996-05-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Magnetospirillum AMB-1
; FEATURE:
; NAME/KEY: gene
; LOCATION: (217)..(702)
; OTHER INFORMATION: Predicted region encoding membrane bound portion
; OTHER INFORMATION: of the mps gene product. "n" is any of a, t, c or
; OTHER INFORMATION: g.
US-08-973-275-2
```

```
Query Match 3.2%; Score 32.6; DB 4; Length 954;
Best Local Similarity 24.0%; Pred. No. 0.53;
Matches 71; Conservative 53; Mismatches 172; Indels 0; Gaps 0;

Qy 545 acagaattattctggactgtgtttcatctctctgtggtgagtgagtggttacttca 604
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 785 ACNSWRTCDATDATNSWARYTFTGNARTCYTGNCGNTNARNCKNARYTGYTCNGCN 726

Qy 605 ggtaaatcagcattgttactgtgttattcaagtaaatgttataattacttccgtgata 664
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 725 GCRTCYITNGCCTTYTCNGCNSWNCCKCANARDATNSWNGRCANCCYTCNGNSWDATN 666

Qy 665 cctctgtgggaagcaggtttttgatacatgcagcttctctgtgattgattgattgattga 724
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 665 ACNSWRTADATNGCTGYTCNARCATNARNACNAGTGTTCNGTNGTNGCNGNARNGCDATNGCN 606

Qy 725 actcaagagaactttgtctcatgtgattctttttaaaccgagtgagtagaactgtctgatg 784
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 605 CCNCCNSWNCNCCYTCNCCDATDATNACNSWNCNARNGNACNCKNACRTTNARRCAN 546

Qy 785 ctctcaataaagttggtcttgcacagagcgttagtctgctcttcttcttcttctctcc 840
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 545 GYTCDATNSWNCNCGCDATNGCYTCNGCNCNCKNGCYTCNGCNCNCKNGCYTCNGCNCNCC 490

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
```





us-09-101-423a-1.rni

Mon Nov 6 10:14:07 2000

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,794

FILING DATE: 24-FEB-1997  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.

REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: PPL96-04

TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1980 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
US-08-804-794-1

Query Match 3.0%; Score 30.6; DB 4; Length 1980;

Best Local Similarity 48.6%; Pred. No. 3.7; Mismatches 0; Gaps 0;  
Matches 84; Conservative

QY 518 gagctgtgacccattctcaagatacagaattattcttggaactgtgttttcagctc 577  
|||||

Db 1647 GAGTGACTAATAATATGTTTGTAGTAAACATATTTATATATAGGATTCATTCAT 1588  
|||||

QY 578 ctgtgtgagggagtgagtttacttcaggttaatacagcattgtcttactgttggtattca 637  
|||||

Db 1587 GACCTCGACCGTGTGTTCTTACTTACTTGTGGAACATATATATTTGCGTGAACATAAT 1528  
|||||

QY 638 gtaaatgcttaaatattctcgtgataacacctctgtggagcaggtttttgata 690  
|||||

Db 1527 ATAGTCTTCATCCATTTGGATTAAATTTGATTAGTCCATTTGGATA 1475  
|||||

RESULT 9

US-08-906-769-130  
Sequence 130, Application US/08906769

Patent No. 6077687;  
GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.

APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.

APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.

APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID

TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh

STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver

STATE: Colorado  
COUNTRY: USA

ZIP: 80203  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,769

FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075

FILING DATE: 24-APR-1996

APPLICANT: Malboobi, Mohammad A.  
TITLE OF INVENTION: Phosphate-Deficiency Inducible Promoter

NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive

CITY: Lexington  
STATE: Massachusetts

COUNTRY: USA  
ZIP: 02173

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804,794  
FILING DATE: 24-FEB-1997

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: PPL96-04  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 780 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
US-08-804-794-2

Query Match 3.0%; Score 30.6; DB 4; Length 780;

Best Local Similarity 48.6%; Pred. No. 2.2; Mismatches 0; Gaps 0;  
Matches 84; Conservative

QY 518 gagctgtgacccattctcaagatacagaattattcttggaactgtgttttcagctc 577  
|||||

Db 447 GAGTGACTAATAATATGTTTGTAGTAAACATATTTATATATAGGATTCATTCAT 388  
|||||

QY 578 ctgtgtgagggagtgagtttacttcaggttaatacagcattgtcttactgttggtattca 637  
|||||

Db 387 GACCTCGACCGTGTGTTCTTACTTACTTGTGGAACATATATATTTGCGTGAACATAAT 328  
|||||

QY 638 gtaaatgcttaaatattctcgtgataacacctctgtggagcaggtttttgata 690  
|||||

Db 327 ATAGTCTTCATCCATTTGGATTAAATTTGATTAGTCCATTTGGATA 275  
|||||

RESULT 8

US-08-804-794-1/c  
Sequence 1, Application US/08804794

Patent No. 5922564  
GENERAL INFORMATION:

APPLICANT: Lefebvre, Daniel D.  
APPLICANT: Malboobi, Mohammad A.

TITLE OF INVENTION: Phosphate-Deficiency Inducible Promoter  
NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive  
CITY: Lexington

STATE: Massachusetts  
COUNTRY: USA

ZIP: 02173  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS



US-07-807-043B-5 ; MOLECULE TYPE: genomic DNA

	Query Match	2.9%	Score 30.4	DB 1	Length 4698
	Best Local Similarity	57.3%	Pred. No. 7.1		
	Matches 55	Conservative 0	Mismatches 41	Indels 0	Gaps 0
Qy	4	cttgggtgctctatgtcttggctctccctctctccagtcaccattagccataaccattctt	63		
Db	1267	CCATTTTCGGGCGCTTTTCTTTCTGTGTCCTCCCTCCGCCCTACCTTTACCAAGCTT	1326		
Qy	64	gacagactctgggacagtcctccctctgctctcctgtt	99		
Db	1327	TGCTCTCCCTGTGTCCTCCCTCCGCCCTTTTGACGCTTTT	1362		

```

RESULT 12
US-08-299-849B-5
; Sequence 5, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:
; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
; APPLICANT: Leth, Bernard; Szkora, Jean-Pierre; De Smet, Charles;
; APPLICANT: Chomez, Patrick
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
; TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; APPLICATION NUMBER: 9-JULY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-May-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5612201man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

; TOPOLOGY: linear  
 ; MOLECULE TYPE: genomic DNA  
 US-08-299-849B-5

	Query Match	2.98;	Score 30.4;	DB 1;	Length 4698;
	Best Local Similarity	57.33;	Pred. No. 7.1;		
	Matches 55;	Conservative 0;	Mismatches 41;	Indels 0;	Gaps 0;
Qy	4	ccttggtgctatgtcttgcctctccctctctccagtcctccattgaagccataaccatttt	63		
Db	1267	CTTTTTCGCGCTTTTCTTCCTGCTCCCTCCCGCTCCCTATTTCACCGATT	1326		
Qy	64	gacagactctggggacagtcctccctctgctctcctgtt	99		
Db	1327	TGCTCTCCCTGCTCCCTCCCGCTTTTGACCGTTT	1362		

RESULT 13  
 US-08-142-368A-5  
 Sequence 5, Application US/08142368A  
 Patent No. 5925729  
 GENERAL INFORMATION:  
 APPLICANT: Bocq-Falleur, Thierry; Van der Bruggen, I  
 APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; I  
 APPLICANT: Lurquin, Christophe; Chomez, Patrick; T  
 TITLE OF INVENTION: Tumor Rejection Antigen Precu  
 TITLE OF INVENTION: Rejection Antigens and Uses Th  
 NUMBER OF SEQUENCES: 26  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Felfe & Lynch  
 STREET: 805 Third Avenue  
 CITY: New York City  
 STATE: New York  
 ZIP: 10022  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
 COMPUTER: IBM  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: Wordperfect  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/142.368A  
 FILING DATE: 02-MAY-1994  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: PC7/US92/04354  
 FILING DATE: 22-MAY-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/807,043  
 FILING DATE: 12-DECEMBER-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/764,364  
 FILING DATE: 23-SEPTEMBER-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/728,838  
 FILING DATE: 9-JULY-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/705,702  
 FILING DATE: 23-MAY-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Hanson, No. 5925729man D.  
 REGISTRATION NUMBER: 30,946  
 REFERENCE/DOCKET NUMBER: LUD 5253.4-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 688-9200  
 TELEFAX: (212) 838-3884  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4698 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: genomic DNA

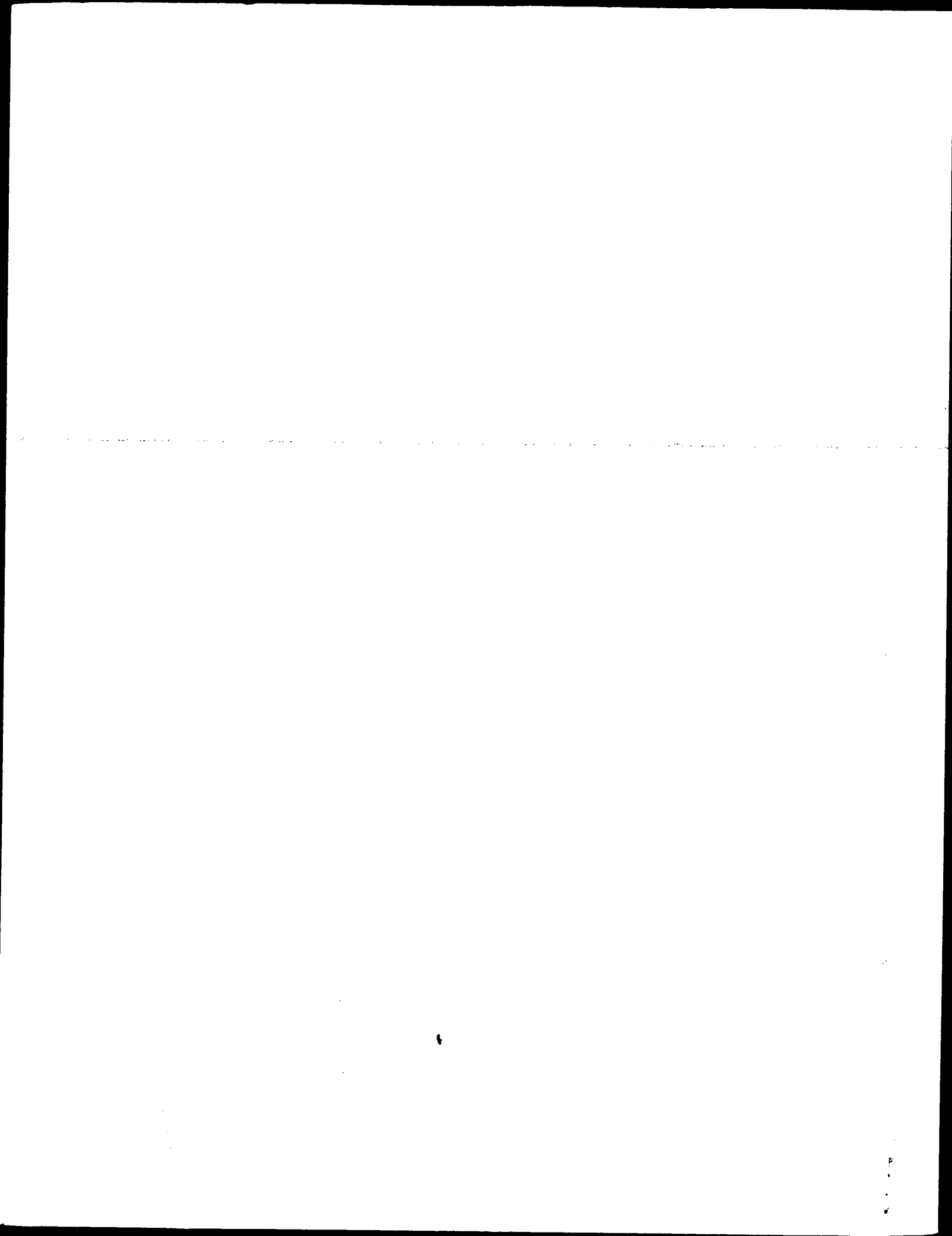


us-09-101-423a-1.rni

Mon Nov 6 10:14:07 2000

Db 910 GGTGACA 916  
|||||

Search completed: November 4, 2000, 00:09:28  
Job time: 21972 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2000, 14:51:06 ; Search time 3273.53 Seconds  
(without alignments)  
1951.056 Million cell updates/sec

Title: US-09-101-423a-1  
Perfect score: 1033  
Sequence: 1 ctctctgggtgctctatgtc.....gtagacataactcccatgt 1033

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues 14379728  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*

44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estom.\*  
68: em\_estov1.\*  
69: em\_estov2.\*  
70: em\_estp11.\*  
71: em\_estp12.\*  
72: em\_estp13.\*  
73: em\_estp14.\*  
74: em\_estp15.\*  
75: em\_estro1.\*  
76: em\_estro2.\*  
77: em\_estro3.\*  
78: em\_estro4.\*  
79: em\_estro5.\*  
80: em\_estro6.\*  
81: em\_estro7.\*  
82: em\_estro8.\*  
83: em\_estro9.\*  
84: em\_estro10.\*  
85: em\_estro11.\*  
86: em\_estro12.\*  
87: em\_estro13.\*  
88: gb\_gss1.\*  
89: gb\_gss2.\*  
90: gb\_gss3.\*  
91: gb\_gss4.\*  
92: em\_gss1.\*  
93: em\_gss2.\*  
94: em\_gss3.\*  
95: em\_gss4.\*  
96: gb\_gss5.\*  
97: gb\_gss6.\*  
98: gb\_gss7.\*  
99: gb\_gss8.\*  
100: gb\_gss9.\*  
101: em\_gss5.\*  
102: em\_gss6.\*  
103: em\_gss7.\*  
104: em\_gss8.\*  
105: em\_gss9.\*  
106: em\_gss10.\*  
107: em\_gss11.\*  
108: gb\_gss10.\*  
109: gb\_gss11.\*  
110: em\_gss12.\*  
111: gb\_gss12.\*  
112: gb\_gss13.\*  
113: gb\_gss14.\*  
114: gb\_gss15.\*  
115: gb\_gss16.\*  
116: gb\_gss17.\*

117: gb-gss18:\*  
 118: gb-gss19:\*  
 119: em-gss13:\*  
 120: gb-gss20:\*  
 121: gb-gss21:\*  
 122: gb-gss22:\*  
 123: gb-gss23:\*  
 124: gb-gss24:\*  
 125: em-gss14:\*  
 126: em-gss15:\*  
 127: em-gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	140.4	13.6	520	117	AZ236681	RPCI-23-8
C 2	124.8	12.1	589	112	AQ224576	RPCI-23-2
C 3	112.4	10.9	579	113	AZ000233	RPCI-23-3
C 4	111.6	10.8	506	118	AZ293934	RPCI-23-1
C 5	100	9.7	632	117	AZ256093	RPCI-23-1
C 6	98.4	9.5	528	115	AZ122046	RPCI-23-1
C 7	96.6	9.4	622	115	AZ093299	RPCI-23-4
C 8	94.2	9.1	364	115	AZ121928	RPCI-23-1
C 9	92	8.9	537	115	AZ110925	RPCI-23-4
C 10	91.6	8.9	475	33	BE106967	UI-R-BT1
C 11	91.6	8.9	620	117	AZ279461	RPCI-23-1
C 12	89.4	8.7	519	118	AZ296888	RPCI-23-1
C 13	86.6	8.4	714	117	AZ282692	RPCI-23-1
C 14	85.2	8.2	449	117	AZ257258	RPCI-23-1
C 15	84.4	8.2	688	113	AZ010006	RPCI-23-2
C 16	84.2	8.2	432	114	AZ030591	RPCI-23-3
C 17	83.8	8.1	527	118	AZ300022	RPCI-23-1
C 18	82.4	8.0	643	115	AZ121474	RPCI-23-1
C 19	80.4	7.8	306	36	D18871	MUSGS01053
C 20	80.2	7.8	497	114	AZ020522	RPCI-23-2
C 21	76	7.4	297	16	AV118376	AV118376
C 22	74.2	7.2	504	114	AZ030252	RPCI-23-3
C 23	73	7.1	621	113	AZ015292	RPCI-23-2
C 24	68.2	6.6	527	19	AW060730	UI-M-BH1
C 25	67.6	6.5	463	115	AZ105847	RPCI-23-3
C 26	67.4	6.5	473	114	AZ080734	RPCI-23-3
C 27	65.4	6.3	266	29	BZ278155	BZ278155
C 28	65.2	6.3	311	30	BZ395634	BZ395634
C 29	63.6	6.2	442	114	AZ032845	RPCI-23-3
C 30	63.6	6.2	467	114	AZ033375	RPCI-23-3
C 31	63.6	6.2	607	114	AZ071951	RPCI-23-3
C 32	63.4	6.1	551	113	AZ000638	RPCI-23-3
C 33	63	6.1	503	114	AZ026762	RPCI-23-3
C 34	63	6.1	536	117	AZ283149	RPCI-23-1
C 35	62.6	6.1	576	117	AZ239870	RPCI-23-8
C 36	61.8	6.0	255	18	AV381235	AV381235
C 37	61.2	5.9	673	117	AZ230129	RPCI-23-8
C 38	60.8	5.9	523	116	AZ223837	RPCI-23-5
C 39	59.8	5.8	552	113	AZ019781	RPCI-23-2
C 40	58.8	5.7	158	114	AZ042155	RPCI-23-3
C 41	58.8	5.7	563	117	AZ249546	RPCI-23-7
C 42	58.6	5.7	411	113	AQ982468	RPCI-23-2
C 43	58.4	5.7	173	115	AZ089053	RPCI-23-6
C 44	58	5.6	440	7	AA967133	ua40f11.r
C 45	58	5.6	574	114	AZ043744	RPCI-23-3

## ALIGNMENTS

RESULT 1  
 AZ236681/c

LOCUS  
 DEFINITION RPCI-23-37D8.TV RPCI-23 Mus musculus genomic clone RPCI-23-87D8, DNA sequence.  
 ACCESSION AZ236681  
 VERSION AZ236681.1 GI:8544727  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 520)  
 AUTHORS Zhao S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.  
 TITLE Mouse BAC End Sequences from Library RPCI-23  
 JOURNAL Unpublished (1999)  
 COMMENT Other GSSs: RPCI-23-87D8.TJ  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 87 row: D column: 8  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers  
 1..520  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-87D8"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site: 1; ECoRI; Site 2: ECoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of ECoRI and ECoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the ECoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 156 a 115 c 113 g 136 t  
 ORIGIN

Query Match 13.6%; Score 140.4; DB 117; Length 520;  
 Best Local Similarity 77.8%; Pred. No. 1.6e-31;  
 Matches 207; Conservative 0; Mismatches 56; Indels 3; Gaps 3;  
 QY 605 ggtaatacagattgcttactgtgttgcattcaagtaaatgtaattatcttgatata 664  
 Db 383 GATTACTCATATTGCTGCTGCTGTTTACCCATTAATGTTTAAACATATCCTTTATGTT 324  
 QY 665 cctctgtggg-aagcagggtttttgatacatgcagcttgccttgccttgattg-atactgctt 722  
 Db 323 CCTCTATGGGAAACATGTTTTTGGCATGTGCAGCTTCTCCTGTGATTGATGCAGCTT 264  
 QY 723 gaactcaagaagactttctctatgtatctttcttaaccgagtgagtagaactgtctga 782  
 Db 263 GRACATCATGAGGACTTTACTATGCGACTCTCTTAACCGGTACAAATTTGCTGA 204  
 QY 783 tgccttcaatgaattggctcttgcagcagcagcttagctgctctgtttatctctccat 842  
 Db 203 TGCTCTGAATAAATGCTACTGCTGCTGAGACATTAGTTCACCGTGTATTATCAGCTCAT 144







BACAPC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 11 row: N column: 17  
Seq primer: T7  
Class: EAC ends.

# FEATURES

source  
1. 528  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-11n17"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site.1:  
ECORI; Site.2: EORI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EORI and EORI Methyase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
ECORI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 153 a 107 c 94 g 174 t  
ORIGIN

Query Match 9.5%; Score 98.4; DB 115; Length 528;  
Best Local Similarity 65.6%; Pred. No. 8.4e-19;  
Matches 231; Conservative 0; Mismatches 96; Indels 25; Gaps 5;

QY 534 ttcttcagatcacagaattattcttgactgtgttttcattgtctcc-----ttgtgg 584  
DB 35 TTCTTCAACATGCAATTCCTCTGGAATGCTTCTGTATTCCTCCAGGCTAGCAT 94  
QY 585 atgggagtgagttacttcacggtaataacgacattgttctgtattcaagtaaatg 644  
DB 94 ATCAGACTGAGTTTACTTCTGATTTATTTACTTATTTACTTGTGTGTGTTCAATGAAGAAG 154  
QY 645 cttaaatctctggatatacctctgtgggaagcaggtttttgatacatgcagctgtgccc 704  
DB 155 TTTAACTATCTTCTTCTATGCCACTATGCGCAAGATGTTT-----TTGTCC 201  
QY 705 ttgtgattg-atactgttgtaactcaagagacattgtctcatgtgattctt-aaacg 762  
DB 202 TTGTGATTGATACAGCTTGAACCTTGTGAGAACTTTACTCATGTGACCTTTTAAAGGCC 261  
QY 763 atggagtagaactgtctgctcctaataaagttgctcttcgacgagacgttagtct 822  
DB 262 TTTTAGTATAAATGCTTAAGGTACTTAAAGAGGTGGTACTACTATGAGACATTAGTCC 321  
QY 823 gtccctgtttatctgtccattcttcc-gctccacagcgcctctacagcactaa 873  
DB 322 ACCTTCTTATCAACTCACTTCTCCAGATCCCAACCTCTATAGCTGTGA 373

# RESULT 7

AZ093299/c 622 bp DNA GSS 08-MAY-2000  
LOCUS  
DEFINITION  
RPCI-23-473H15.TV RPCI-23 Mus musculus genomic clone RPCI-23-473H15  
DNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

/clone="RPCI-23-136C21"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site.1:  
ECORI; Site.2: EORI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EORI and EORI Methyase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
ECORI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 184 a 109 c 132 g 207 t  
ORIGIN

Query Match 9.7%; Score 100; DB 117; Length 632;  
Best Local Similarity 66.9%; Pred. No. 2.9e-19;  
Matches 241; Conservative 0; Mismatches 95; Indels 24; Gaps 6;

QY 534 ttcttcagatcacagaattattcttgactgtgttttcattgtctcc-----ttgtgg 584  
DB 374 TTCTTCAACATGCAATTCCTCTGGAATGCTTCTGTATTCCTCCAGGCTAGCAT 315  
QY 585 atgggagtgagttacttcacggtaataacgacattgttctgtattcaagtaaatg 644  
DB 314 ATAGGAATGAGTTTACTTCAATAAATCAATATTGTTGCTGTGTTAAATCAATTAATG 255  
QY 645 cttaaatctcc-----tgatatacctctgtgggaagcaggtttttgatacatgcagc 698  
DB 254 TTTAAGCTATCTTCTTATGGAAGCCTGATGGAAGAAACATGTTTGGCCACGACG 195  
QY 699 ttgtcttgattg-atactgttgtaactcaagacattgttctcatgtg----attctt 754  
DB 194 TTTCTTTGTCGCTGTATACAGCTGACCCATGAGATTTTACTTCTGATTTATTTT 135  
QY 755 cttaaccagatgagtt-----agaaactgtctgactctctcctaataaagttgctcttcacg 810  
DB 134 CTTAACTTCAAGTATAAATAAATATATCTGAGCCTGCAAGTAGTGTGCTAGTAGTG 75  
QY 811 agacttagtctgtctttatctgtccattct-tccgtccacagcgcctctacagca 869  
DB 74 AAACCTTGTCTTACCTGTTTATGTACTCTCAGCTCTCAGGTCCTCTAGAGTA 15

# RESULT 6

AZ122046 528 bp DNA GSS 12-MAY-2000  
LOCUS  
DEFINITION  
RPCI-23-11n17.TV RPCI-23 Mus musculus genomic clone RPCI-23-11n17,  
DNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

## COMMENT

Other\_GSSs: RPCI-23-473H15.TJ  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 473 row: H column: 15  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
 1. .622  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-473H15"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1: EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 189 a 145 c 116 g 172 t

BASE COUNT  
ORIGIN

Query Match 9.4%; Score 96.6; DB 115; Length 622;  
 Best Local Similarity 67.7%; Pred. No. 3.1e-18;  
 Matches 239; Conservative 0; Mismatches 79; Indels 35; Gaps 6;  
 QY 534 ttcttaagatcacagaattattcttggaactgtgttttcactctctt-----gtgg 584  
 Db 394 TTCTTCAGCATACAGAAATGTTCTGGAGTATGCTTTGATGCTGCCAGGTGCAGTGT 335  
 QY 585 atgggagtgagttacttcaggttaatacagcactgcttactgttggtattcaagtaaatg 644  
 Db 334 AGAGAGAGTGAGTTACTTCAGATGGCTCATTTATTCACGCTGTGTCTATTCAATTAATG 275  
 QY 645 cttaaatatctctggatatacctctgtgggaagca---ggtttttgatcacatgcagcttg 701  
 Db 274 TTAAACTATCTCTTATATGCTTTATGGAAGACATGGGTTTGTGGATGCAGTTTG 215  
 QY 702 tctgtgtatg-atactgctgaactcaagagaactttgtctcatgtatctttcttaac 760  
 Db 214 TCCATGATGTTGACAAGAAATGTTCTCAAGATAAGCTTACTCAT-----168  
 QY 761 cgatggagtagaactgtctgagtctctcaa-taaagttggctctctgcacagagagttag 819  
 Db 169 ----GAGAAATGAATTCCTGAGGCTCTGAATTAAGTTGCCCTACTGCATTAGACTTCAG 115  
 QY 820 tctgtctgtttactctctccattct-tccgctcccacggcctctacagcact 871  
 Db 114 TCCCCCTATTATTGCTCCATTCTCTCAGGTCCTGCCACCTCTCTAGCACT 62

## RESULT 8

AZ121928  
 LOCUS AZ121928 364 bp DNA GSS 12-MAY-2000  
 DEFINITION RPCI-23-1J15.TV RPCI-23 Mus musculus genomic clone RPCI-23-1J15,  
 DNA sequence.  
 ACCESSION AZ121928  
 VERSION AZ121928.1 GI:7789322

KEYWORDS  
SOURCE  
ORGANISM

GSS.  
 house mouse.  
 Mus musculus

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 364)  
 Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Aklnret,  
 B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
 and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 Other\_GSSs: RPCI-23-1J15.TJ  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 1 row: J column: 15  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
 1. .364  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-1J15"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1: EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 101 a 67 c 69 g 127 t

BASE COUNT  
ORIGIN

Query Match 9.1%; Score 94.2; DB 115; Length 364;  
 Best Local Similarity 71.7%; Pred. No. 1.4e-17;  
 Matches 137; Conservative 0; Mismatches 53; Indels 1; Gaps 1;  
 QY 675 aagcaggttttgcatacatgcagctgtctgtctgtgattg-atactgctgaactcaag 733  
 Db 67 AACATGTTTACGTACATGTGCTGTCTTCTGTGATTGAATACAGCTGAACACATGAG 126  
 QY 734 aacttgctctatgtatctttcttaaccgatgagtagaactgtctgtctctcaata 793  
 Db 127 AACTTTACTCAGTGCAGCTTCTTAATCTCCAGACTATAATGTTCTGTGCTTTAAATA 186  
 QY 794 aagttgctctgcacagacagcttactgtctgtcttttatctgtccatttccgcctcc 853  
 Db 187 AAGTTGGATTGGCAAGACTTTAGTTTCGGCTCATCTTTAGGCTCTGCTCTCCAGGTT 246  
 QY 854 cagcgcctctta 864  
 Db 247 CATGCTATCA 257

## RESULT 9

## LOCUS

## DEFINITION

AZ110925 537 bp DNA GSS 09-MAY-2000  
 RPCI-23-469013.TV RPCI-23 Mus musculus genomic clone RPCI-23-469013





Mon Nov 6 10:14:08 2000

266 TAAAGTTTGAATCTCTATATACCTCTATGG--AATAGTTTTCCTCGTGGCAGC 209  
 699 ttgccccttgatg-atactgttgaactcaagaagaactttgtctcatgtgatctttctt 757  
 208 TTATCTCTGTGATGATATAGATGGAATTTATACATATGTTCTCTGTGACTATTCTT 149  
 758 aaccgatggatgagaactctctgatctctcacaataaaagtggcttgcacagacgtt 817  
 148 AACCATCAGAGTATAAATTTATGA-CGTGTCATAAAGTTGGCTTT-----GCATAAGAA 94  
 818 agtctgctctgtttatctctctcattctcc 848  
 93 ATATTGGCTCATTTATTGGCTCAATCTCTCC 63

RESULT 13  
 AZ282692 714 bp DNA GSS 27-JUL-2000  
 LOCUS AZ282692.1 GI:9524478  
 DEFINITION RPI-23-138L11.TJ RPI-23 Mus musculus genomic clone RPI-23-138L11  
 , DNA sequence.

ACCESSION AZ282692  
 VERSION AZ282692.1  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 714)  
 Zhaoh.S., Nierman.W., Feldblyum.T., Malek.J., Shatsman.S., Akinret  
 and and Fraser.C.M.  
 Mouse BAC End Sequences from Library RPI-23  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhaoh@tigr.org  
 Clones are derived from the mouse BAC library RPI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
 or from Resea ch Genetics (info@resgen.com). BAC end page:  
 http://www.tigr.org/tdb/bac-ends/mouse/bac\_end\_intro.html  
 Plate: 138 row: L column: 11  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 1..714  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPI-23-138L11"  
 /clone\_lib="RPI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
 EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBACe3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 190 a 150 c 164 g 210 t  
 ORIGIN

Query Match 8.4%; Score 86.6; DB 117; Length 714;  
 Best Local Similarity 61.2%; Pred. No. 3.5e-15;  
 Matches 260; Conservative 0; Mismatches 129; Indels 36; Gaps 6;

QY 439 ttagggaaccactgtgttgaacatgtccaagaatacacagtcatgtgaggagataccca 498  
 DB 304 TTCTGTAGCCCAAGTGTCTAGAACATAGTCTTAGAATGAGTCATGCAAGGAGATGCTG 363  
 QY 499 atgcgtcaggagaaaaac-----gagagctctgtgacctccattcttcaagata 545  
 DB 364 AGTGTCTGTGAGAAAACGCCGAGAAAACAAAGAGTCTTGTGACCCGAGTTTCTTCAAACCG 423  
 QY 546 cagaattattcttgactgtt--tttcatgctctgtgagtgaggagtgacttacttc 603  
 DB 424 TGGAGCTGTCTTGGACTGTGTGTTCCTCCAGGTAGCAGATAGAGTGAATTACTTTC 483  
 QY 604 aggttaactcagcattgttactgttgattcaagttaaatgcttaaatatactctggatat 663  
 DB 484 CCATTACTACTA-----TGCTGGGTCAATTCAGCTAAATGTT-----TGT 524  
 QY 664 acctgtgtgggaagcaggtttttgtatatacatgacagcttgcctgtgactgatactgcttg 723  
 DB 525 GCCTTACACTAAACATGCTTTGCCACATGTGGCTCGCTCTGTGATTG-TATAGCTTG 583  
 QY 724 aactcaagagaactttgctcatgtgatctttcttaaccgatgagtagaactgtctgat 783  
 DB 584 AACTCATGACAACCTTATTTCATGTGAATTCCTTACCCCTCAGAGTATCAGTTGTCCTCAAT 643  
 QY 784 gcttcaataaagtgtgctcttgcacgacagcgttagctgtctctgttcttatctgctccatt 843  
 DB 644 GCCTGTGAAT-AGTTAGCTATTGATGACACATTTAATATGCCCATTTATTAGTCCCACT 702  
 QY 844 ctccc 848  
 DB 703 CTGCTC 707

RESULT 14  
 AZ257258 449 bp DNA GSS 26-JUL-2000  
 LOCUS RPI-23-136G19.TV RPI-23 Mus musculus genomic clone RPI-23-136G19  
 DEFINITION , DNA sequence.

ACCESSION AZ257258  
 VERSION AZ257258.1  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 449)  
 Zhaoh.S., Nierman.W., Feldblyum.T., Malek.J., Shatsman.S., Akinret  
 and and Fraser.C.M.  
 Mouse BAC End Sequences from Library RPI-23  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhaoh@tigr.org  
 Clones are derived from the mouse BAC library RPI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
 or from Resea ch Genetics (info@resgen.com). BAC end page:  
 http://www.tigr.org/tdb/bac-ends/mouse/bac\_end\_intro.html  
 Plate: 136 row: G column: 19  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 source  
 1..449  
 /organism="Mus musculus"  
 /strain="C57BL/6J"

Query Match 8.4%; Score 86.6; DB 117; Length 714;  
 Best Local Similarity 61.2%; Pred. No. 3.5e-15;  
 Matches 260; Conservative 0; Mismatches 129; Indels 36; Gaps 6;

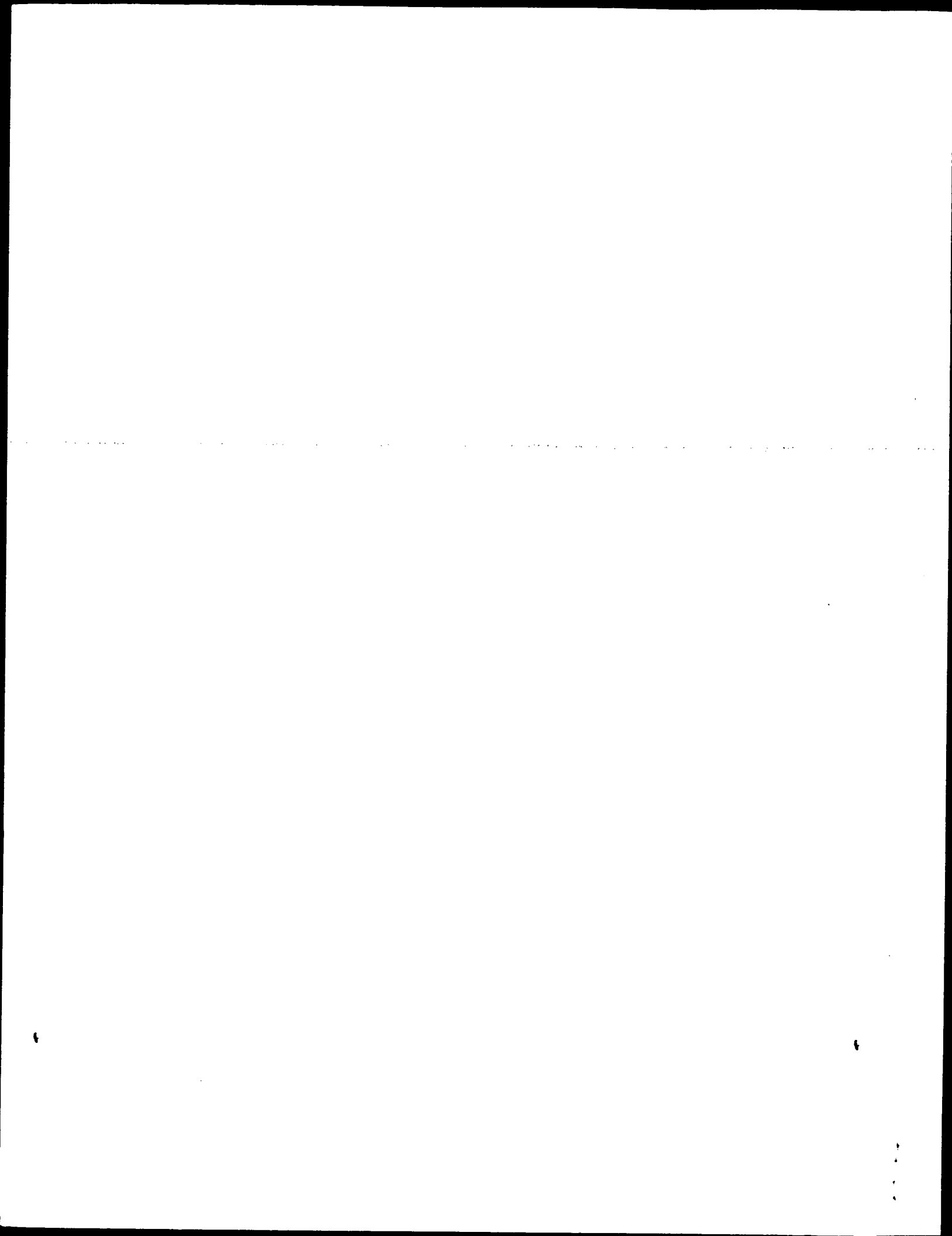




Mon Nov 6 10:14:08 2000

us-09-101-423a-1.rst

Page 11



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:36:43 ; Search time 7111.01 Seconds  
(without alignments)  
649.823 Million cell updates/sec

Title: US-09-101-423a-2  
Perfect score: 1058  
Sequence: 1 attgctgtgagcctattagc.....acacattcgacgcttggtc 1058

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1033670 seqs, 2183789903 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:

1: gb\_bal.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pr1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: em\_fun.\*  
13: em\_hum1.\*  
14: em\_hum2.\*  
15: em\_in.\*  
16: em\_om.\*  
17: em\_or.\*  
18: em\_ov.\*  
19: em\_pat.\*  
20: em\_ph.\*  
21: em\_pl.\*  
22: em\_ro.\*  
23: em\_sts.\*  
24: em\_sy.\*  
25: em\_un.\*  
26: em\_vi.\*  
27: gb\_htg1.\*  
28: gb\_htg2.\*  
29: gb\_in1.\*  
30: gb\_in2.\*  
31: em\_bal.\*  
32: em\_ba2.\*  
33: em\_hum3.\*  
34: em\_hum4.\*  
35: gb\_pr4.\*  
36: gb\_htg3.\*  
37: gb\_htg4.\*  
38: gb\_htg5.\*  
39: gb\_htg6.\*  
40: gb\_htg7.\*  
41: em\_htg1.\*  
42: em\_htg2.\*  
43: em\_htg3.\*

44: em\_hum5.\*  
45: gb\_pl3.\*  
46: gb\_pr5.\*  
47: gb\_htg8.\*  
48: gb\_htg9.\*  
49: gb\_htg10.\*  
50: gb\_htg11.\*  
51: gb\_htg12.\*  
52: gb\_htg13.\*  
53: gb\_htg14.\*  
54: gb\_in3.\*  
55: gb\_htg15.\*  
56: gb\_htg16.\*  
57: gb\_htg17.\*  
58: em\_htg4.\*  
59: em\_htg5.\*  
60: em\_htg6.\*  
61: em\_htg7.\*  
62: em\_hum6.\*  
63: gb\_htg18.\*  
64: gb\_htg19.\*  
65: gb\_ba3.\*  
66: em\_htg8.\*  
67: em\_htg9.\*  
68: em\_htg10.\*  
69: em\_htg11.\*  
70: em\_htg12.\*  
71: em\_htg13.\*  
72: em\_htg14.\*  
73: em\_htg15.\*  
74: em\_htg16.\*  
75: em\_htg17.\*  
76: em\_htg18.\*  
77: em\_htg19.\*  
78: em\_htg20.\*  
79: em\_htg21.\*  
80: em\_htg22.\*  
81: em\_htg23.\*  
82: gb\_pr6.\*  
83: gb\_pr7.\*  
84: gb\_htg20.\*  
85: gb\_htg21.\*  
86: gb\_htg22.\*  
87: gb\_htg23.\*  
88: gb\_ro.\*  
89: gb\_sts1.\*  
90: gb\_sts2.\*  
91: gb\_sy.\*  
92: gb\_un.\*  
93: gb\_vil.\*  
94: gb\_vil2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1058	100.0	1058	5	A64059 Sequence 2
C	2	25	2.4	152402	48 AC022658
3	25	2.4	163317	35	CNS01DX0
C	4	25	2.4	195713	87 CNS05TBP
5	22	2.1	154284	88	AC005992
6	21	2.0	945	90	G65741
7	21	2.0	37321	9	AC005498
C	8	21	2.0	116917	9 AC003024
9	21	2.0	131047	27	AC002421
10	21	2.0	136410	49	AC023308
11	21	2.0	152355	10	AC005668
C	12	21	2.0	167873	40 AC018652

13	C	14	2.0	172091	27	AC007962	Homo sapi	AC007962 Homo sapi	Qy	121	tgtggtcagcgcagacatttagggatgtgtatggagacaggggtcggggaagaagagaag	180
15		16	1.9	3907	82	HSP0071	X81889 H.sapiens m	X81889 H.sapiens m	Db	121	TGTGTCAGCAGCCAGCAATTTAGGGATGTGATGGACAGGGTCCGGGAAAGAGGAGAAG	180
17	C	17	2.0	1.9	39739	38	AC015123	Drosophill				
18	C	18	2.0	1.9	40957	36	AC011526	Homo sapi	Qy	181	ggttaaaaggaagacagacacgcttaaaagtccaaacagctccaggagactatgttagaata	240
19	C	19	2.0	1.9	61656	63	AC007690	Homo sapi	Db	181	GGTAAAGGAAAGACAGACACGCTTAAAGTCCAAACAGCTCCAGGAGACTATCTGTAGAAATA	240
20	C	20	2.0	1.9	73568	50	AC025967	Homo sapi				
21	C	21	2.0	1.9	77130	29	AC007145	Drosophill	Qy	241	acatcagaccatgagggagaattgatatactgttttcaattgggtatcgccaagggaact	300
22	C	22	2.0	1.9	82263	10	AC007999	Homo sapi	Db	241	ACATCAGACCATGAGGAGAAATGATATCATGTGTTTTCATGGGTATCGCAAGGGAAC	300
23	C	23	2.0	1.9	96933	50	AC025463	Homo sapi				
24	C	24	2.0	1.9	110000	87	LMFLCHR12_08	Continuation (9 of	Qy	301	tccatctgataaaaaataattactgtgccaactaaatccaaatgggaatgccccacaca	360
25		25	2.0	1.9	115173	37	AC012612	Homo sapi	Db	301	TTCCATCTGATTAATAATAATTAAGTCTGGCAGCTTAAATCCAAATTTGGAATGCCCCACACA	360
26		26	2.0	1.9	115199	9	AC003695	Homo sapi	Qy	361	atttatcttcacttcactgctcacatgctcgtacgctgagcgggagacagacagcattccc	420
27	C	27	2.0	1.9	124232	42	AC010750	Drosophill	Db	361	ATTTATCTTCACCTCATGTGCTTACCATATGCCGTGACGTGGCGGACGAAGCATTTCCC	420
28		28	2.0	1.9	133197	85	AL360092	Homo sapi				
29	C	30	2.0	1.9	134221	63	AL136459	Homo sapi	Qy	421	tcccggtctctgataaataagtagtacttggtaaatatttggagacgggagctctggtgacaggga	480
30		31	2.0	1.9	138119	85	AL365209	Homo sapi	Db	421	TCCCGTCTCTGATAAATAGTAGTCTTTGATAATATTGGAGACGGGAGCTCTGGTGACAGGGA	480
31	C	32	2.0	1.9	138675	46	HS1057B20	Human DNA	Qy	481	acagctacaaacgggcctgtttatcatgtttcccgatagagggccctttgacatacagga	540
32	C	33	2.0	1.9	139480	83	HUAC002331	Homo sapi	Db	481	ACACGTACAAACCGGCCTGTTTATCATGTTCCTCCATGAGAGGCCCTCTTTGACGTACAGGA	540
33	C	34	2.0	1.9	148841	11	AC011462	Homo sapi				
34	C	35	2.0	1.9	151104	36	AC010250	Homo sapi	Qy	541	cccaaaacagtcaggatgctgtgaatttcccttccatgaagccttggacacatttagcaa	600
35	C	36	2.0	1.9	151676	57	AC074362	Homo sapi	Db	541	CCCCAAAACAGTCAGGATGCTGTGAATTTCTTCCATGAAGCCTTGTTCACAATTAGCAA	600
36	C	37	2.0	1.9	153467	48	AC022598	Homo sapi				
37	C	38	2.0	1.9	156373	57	AC074105	Oryza sat	Qy	601	ccattggaggaagcagcctcactgtctacacaaagtggcactttccaaagacacacat	660
38	C	39	2.0	1.9	161424	40	AC018367	Homo sapi	Db	601	CCATTGGAGGAGCAGGAGCTGACTGTCTACCAAGTGGCACATTTCCAAAGACACACAT	660
39	C	40	2.0	1.9	161799	9	AC002091	Genomic s				
40	C	41	2.0	1.9	162107	38	AC013768	Homo sapi	Qy	661	atattggaggaagacacatttctggtgactggtgctgtttaaagctgataaaactgctata	720
41	C	42	2.0	1.9	162495	38	AC015647	Homo sapi	Db	661	ATATTGGAGCAAGACATTTTGTGGCTGACTGGTGCTGTGTAAGCTGATAAACTGCTATA	720
42	C	43	2.0	1.9	162495	38	AC015647	Homo sapi	Qy	721	tttatataaacggccttttttgaacacccccactcaaggaagaaaaaacacacatttagggt	780
43	C	44	2.0	1.9	165385	64	AL158042	Homo sapi	Db	721	TTTTATTAACATGGCTTTTCTTTGAACACCCCACTCAAGGAAAAAATAAACACACTTAGGCT	780
44	C	45	2.0	1.9	165415	39	AC016663	Homo sapi				
45	C	45	2.0	1.9	168237	52	AC034193	Homo sapi	Qy	781	gacattatttggagatgaagctctttatagagatgcttaagctttaaacgagacactttaag	840
									Db	781	GACATTATTGGAGATGAAGCTTTTATAGAGATGCTTAAGTTTAAACGAGACTTTTAAAG	840
									Qy	841	ccggctctatccatttaataatggtgtgtccctacaaaggaagaaactgggacagaggtta	900
									Db	841	CCGGCTCTATTTCATTTAATGAATGGTGTGCCCTACAAAGAGAGAAAACTGGACAGAGTA	900
									Qy	901	tgtacacattg-gt-g	

RESULT 2  
AC022658/c  
LOCUS  
DEFINITION Homo sapiens clone RP11-29617, WORKING DRAFT SEQUENCE, 9 unordered pieces.  
ACCESSION AC022658  
VERSION AC022658.3 GI:7249113  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

AC022658 152402 bp DNA HTG  
16-MAR-2000

SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 152402)  
Birren, B., Linton, L., Nusbaum, C., and Lander, E.  
Homo sapiens chromosome, clone RP11-29G17  
Unpublished  
2 (bases 1 to 152402)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,  
Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Castle, A.,  
Chepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,  
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,  
Ferrelira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howard, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Landers, T., Lehotzky, J., Levine, R., Lieu, C., Liu, G., Locke, K.,  
Macdonald, P., Marquitz, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,  
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,  
Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2000 this sequence version replaced gi:6978198.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L4832  
Center clone name: 29\_G17  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 146863 bases at least Q40  
Consensus quality: 149885 bases at least Q30  
Consensus quality: 150971 bases at least Q20  
Insert size: 155000; agarose-fp  
Insert size: 151602; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 5013: contig of 5013 bp in length  
\* 5014 5113: gap of 100 bp  
\* 5114 15407: contig of 10294 bp in length  
\* 15408 15507: gap of 100 bp  
\* 15508 30456: contig of 14949 bp in length  
\* 30457 30556: gap of 100 bp  
\* 30557 45841: contig of 15285 bp in length  
\* 45842 45941: gap of 100 bp  
\* 45942 61650: contig of 15709 bp in length  
\* 61651 61750: gap of 100 bp  
\* 61751 80995: contig of 19245 bp in length  
\* 80996 81095: gap of 100 bp

\* 81096 103199: contig of 22104 bp in length  
\* 103200 103299: gap of 100 bp  
\* 103300 124126: contig of 20827 bp in length  
\* 124127 124226: gap of 100 bp  
\* 124227 152402: contig of 28176 bp in length.  
Location/Qualifiers  
1. .152402  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="RP11-29G17"  
/clone\_lib="RP11-29G17" Human Male BAC"  
1. .5013  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
5114. .15407  
/note="assembly\_fragment"  
15508. .30456  
/note="assembly\_fragment"  
30557. .45841  
/note="assembly\_fragment"  
45942. .61650  
/note="assembly\_fragment"  
61751. .80995  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
81096. .103199  
/note="assembly\_fragment"  
103300. .124126  
/note="assembly\_fragment"  
124227. .152402  
/note="assembly\_fragment"  
47159 a 30831 c 30633 g 42977 t 802 others  
BASE COUNT  
ORIGIN  
Query Match 2.4% Score 25; DB 48; Length 152402;  
Best Local Similarity 100.0%; Pred. No. 0.0037;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 374 ttcatgctgctaccatgctgac 398  
|||||  
Db 40405 TTCATGCTGCTACCATATGCTGAC 40381  
RESULT 3  
CNSOLDX0  
LOCUS  
DEFINITION  
CNSOLDX0 163317 bp DNA PRI 25-MAY-2000  
Human chromosome 14 DNA sequence \*\*\* IN PROGRESS \*\*\* BAC C-2588E21  
of library Caltech-D from chromosome 14 of Homo sapiens (Human),  
complete sequence.  
ACCESSION AL138995  
VERSION AL138995.3 GI:7413823  
KEYWORDS HTG; HTGS-DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 153317)  
Genoscope.  
Direct Submission  
TITLE  
JOURNAL  
COMMENT  
Submitted (23-MAY-2000) to the EMBL/GenBank/DBJ databases  
On Apr 4, 2000 this sequence version replaced gi:7378660.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage, etc. . . even if efforts are made to eliminate these  
contaminating sequences. The following BAC sequence is oriented  
from the T7 to the SP6 end.  
Upstream BAC (overlapping the T7 end) : R-1012E15

Downstream BAC (overlapping the SP6 end) : R-930011

Overall quality chart :

```

Range : bases
0 : 3
1 : 532
10 : 1854
20 : 5241
30 : 13843
40 : 12446
50 : 12756
60 : 26972
70 : 47065
80 : 30271
90 : 12334

```

Percentage of bases with a quality value >= 40 : 86 %.

```

Location/Qualifiers
1. .163317
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="C-2588E21"
/clone_lib="Caltech-D"

```

BASE COUNT 46800 a 31655 c 34205 g 50654 t 3 others

ORIGIN

Query Match 2.4%; Score 25; DB 35; Length 163317;  
Best Local Similarity 100.0%; Pred. No. 0.0037;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 ttcatgctgctaccatgctgac 398  
|||||  
Db 693 TTTCATGCTGCTACCATATGCCCTGAC 717

RESULT 4  
CNS05TBP/c 195713 bp DNA HTG 25-MAY-2000  
LOCUS Homo sapiens chromosome 14 clone R-1012E15, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 9 ordered pieces.  
ACCESSION AL163952  
VERSION AL163952.2 GI:9213040  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 195713)  
Genoscope.  
Direct Submission

Submitted (23-MAY-2000) to the EMBL/GenBank/DBJ databases  
On Jul 15, 2000 this sequence version replaced gi:8218003.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage, etc. . . even if efforts are made to eliminate these  
contaminating sequences. The following BAC sequence is oriented  
from the T7 to the SP6 end.

Upstream BAC (overlapping the T7 end) : C-2588E21  
Downstream BAC (overlapping the SP6 end) : R-813i20 Contigs  
composition :

```

14393 bp contig from 1 to 14393
1243 bp contig from 14494 to 15736
2086 bp contig from 15837 to 17922
1957 bp contig from 18023 to 19979
8034 bp contig from 20080 to 28113
5056 bp contig from 28214 to 33269
87497 bp contig from 33370 to 120866
68538 bp contig from 120967 to 189504

```

6109 bp contig from 189605 to 195713

Overall quality chart :

```

Range : bases
0 : 1227
1 : 1625
10 : 1725
20 : 3285
30 : 13973
40 : 15406
50 : 10828
60 : 19441
70 : 47709
80 : 47803
90 : 32691

```

Percentage of bases with a quality value >= 40 : 88 %.  
\* NONE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

```

1 14393: contig of 14393 bp in length
14394 14493: gap of 100 bp
14434 15736: contig of 1243 bp in length
15737 15836: gap of 100 bp
15837 17922: contig of 2086 bp in length
17923 18022: gap of 100 bp
18023 19979: contig of 1957 bp in length
19980 20079: gap of 100 bp
20080 28113: contig of 8034 bp in length
28114 28213: gap of 100 bp
28214 33269: contig of 5056 bp in length
33270 33369: gap of 100 bp
33370 120866: contig of 87497 bp in length
120867 120966: gap of 100 bp
120967 189504: contig of 68538 bp in length
189505 189604: gap of 100 bp
189605 195713: contig of 6109 bp in length.

```

FEATURES  
source

```

1. 195713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="RPCI-11"
/clone="R-1012E15"

```

BASE COUNT 59397 a 40418 c 39000 g 55988 t 910 others

Query Match 2.4%; Score 25; DB 87; Length 195713;  
Best Local Similarity 100.0%; Pred. No. 0.0037;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 ttcatgctgctaccatgctgac 398  
|||||

Db 38946 TTTCATGCTGCTACCATATGCCCTGAC 38922

RESULT 5

AC005992  
LOCUS AC005992 154284 bp DNA ROD 15-FEB-2000  
DEFINITION Mus musculus chromosome 1q21-23 clone unknown strain 129X1/Svj ES  
cell line RM4, complete sequence.  
ACCESSION AC005992  
VERSION AC005992.15 GI:6143913  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus



to 6,307 and is expected to overlap BAC 37295 to the right by approximately 3 kb. Additional chr 19 map and sequence information are available at: <http://www-bio.llnl.gov/bbrp/genome/genome.html>.

## FEATURES

```

source
1..37321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R31665"
/chromosome="19"
/map="19q13.4 between D19S773 to ZNF134"
/cell_line="5HL2-B"
/clone_lib="LL19NC03 R chromosome 19- specific cosmid library"
/note="Cosmid library LL19NC03 was constructed at LLMML from flow-sorted chromosomes from human-hamster hybrid 5HL2-B which carries chromosome 19 as its only human chromosome."
repeat_region
complement(14..100)
/rpt_family="MER44A"
565..833
/rpt_family="MLT1A1"
complement(898..1254)
/rpt_family="THE1B"
1392..1693
/rpt_family="AluY"
1705..1996
/rpt_family="AluY"
1705..1996
/rpt_family="AluJo"
complement(2002..2256)
/rpt_family="MER7B"
complement(2769..3070)
/rpt_family="AluSg"
3263..3390
/rpt_family="FRAM"
complement(3857..4154)
/rpt_family="AluJo"
complement(4264..4286)
/rpt_family="AT-rich"
4359..4976
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 79.000"
join(<4361..4976,5418..5983)
/note="Hypothetical ZNF-like human protein"
/codon_start=1
/evidence=not experimental
/product="R31665.1"
/protein_id="AAC32422.1"
/db_xref="GI:3445180"
/translation="WESIYVQLPLKQFYDDACMEGITSYGLCSTFEENKRWEDL
FEKQSGHEMFSSKEITHTKTTTKTKFKYKFGKCIHLENIEESLYNHTSKKFSF
KNSWIKHKVYVKGLKFCNECDKFTTHSSSLTVHFRIHTGKPYACEGKAFKQR
QHLAQHRTHTGKLECEKCAKQSEHLQHQRIHTGERPYFFSHHSLTQHQRI
VHSGEPYCEKCAKQFNVLVSHLRHTGKPYCEKCAKFRISQSLATHQRIH
TGKPYCEKCAKQFNVLVSHLRHTGKPYCEKCAKFRISQSLATHQRIH
ERPKYCEKCAKQFNVLVSHLRHTGKPYCEKCAKFRISQSLATHQRIH
P"
5394..5492
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 55.000"
6100..7721
/rpt_family="LIP3/4"
7865..7981
/rpt_family="LINE2"
7991..8102
/rpt_family="MER30"
8278..8351
/rpt_family="MIR"
complement(8471..8563)
/rpt_family="FLAM_A"
9112..10260
/note="DPS similarity to overlapping ESTs:
(9112..9563) AA428231 zv83912.r1 Soares total fetus N62HF8
9w Homo sapiens cDNA clone 760294 5'; Score: 908
Identity: 454/454 (100%).-(9339..9783) AA461597 zv51401.r1
```

```

misc_feature
AA854060, AA634443, AA460308
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 77.000"
complement(10330..10356)
/rpt_family="(CAAAA)n"
complement(10432..10660)
/rpt_family="AluSc"
complement(11208..11393)
/rpt_family="Li"
complement(11395..11696)
/rpt_family="AluSg"
complement(11697..13623)
/rpt_family="LIM1/2"
complement(13611..14587)
/rpt_family="LIM1/2"
complement(14645..14812)
/rpt_family="LIM4"
complement(14902..14942)
/rpt_family="AT-rich"
complement(15756..15792)
/rpt_family="AT-rich"
complement(16911..17305)
/rpt_family="LTR3"
18805..18872
/rpt_family="(CGG)n"
18859..18990
/rpt_family="(CGGG)n"
19692..19751
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 98.000"
20555..20656
/rpt_family="LINE2"
20657..20828
/rpt_family="MER58A"
20874..21643
/rpt_family="LINE2"
complement(21758..22061)
/rpt_family="AluSx"
complement(22198..22485)
/rpt_family="AluJb"
22492..22579
/rpt_family="MER8"
complement(22800..22981)
/rpt_family="AluJo"
23324..23435
/rpt_family="(TA)n"
complement(23437..23602)
/rpt_family="FAM"
23973..24169
/rpt_family="MIR"
complement(24268..24590)
/rpt_family="MER58B"
24594..24830
/rpt_family="MIR"
25850..25835
/rpt_family="MER54"
complement(26055..26228)
/rpt_family="MIR"
complement(27086..27480)
/rpt_family="MLT1B"
27492..27669
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 91.000"
join(27492..27669,27842..27937,28993..29156,30102..30216,
30545..30640,34055..34908)
/note="Hypothetical ZNF-like human product; Strongest
similarity to (X12594) mkr5 gene product [Mus musculus];
```

CDS



```

ZF28_MOUSE_ZINC_FINGER_PROTEIN_ZFP-28 (MKR5 PROTEIN)"
/codon_start=1
/evidence=not_experimental
/product="R31665_2"
/db_xref="GI:3445181"
/translating="MGCHEWLSKNWISFQGLVFGDVAVDFSOEWEWLNPIORNL
YRKVLENTNRLASGLGVSKPDVLSLQGGKEPTVKKRMRACPDLPKAWKIKEL
PLKDFCEKLSQAVITERTSYNLEYSLLGEHWDYDALFETOPGLVTKNLAVDPRQ
OLHPKQNFCKNGIWNENSDLAGSCHVAKPDLVSLLEQEPKPMVKRELTSGLFSGI
YAGKLFKNECKKTQSSSLTHORIHTEKPKYKCEGKAFSDGSGSFARHQRCHT
GKPYCEGCKAFONTSLIHRWYHTEGKPCDIDCGKAFSDHIGLHOHRRHTG
MKYKCDVCHKFSGASSLTHORIHTEKPKYKCEGKAFSDHIGLHOHRRHTG
PFKCEGKAFONTSLIHRWYHTEGKPCDIDCGKAFSDHIGLHOHRRHTG
EKCVCSEKAFONTSLIHRWYHTEGKPCDIDCGKAFSDHIGLHOHRRHTG
27842..27937
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 63.000"
28993..29156
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 100.000"
29431..29722
/rpt_family="Alu"
30102..30216
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 100.000"

Query Match      2.0%  Score 21;  DB 9;  Length 37321;
Best Local Similarity 100.0%;  Pred. No. 0.97;
Matches 21;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 636 gtggcactttccaagagcac 656
|||||
DB 13640 GTGGCACTTCCAAAGACAC 13660

RESULT 8
AC003024/c 116917 bp DNA PRI 16-SEP-1998
LOCUS Human Chromosome 15q26.1 PAC clone pDJ41616, complete sequence.
DEFINITION AC003024
ACCESSION AC003024
VERSION AC003024.1 GI:3608160
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 116917)
Evans,G.A., Athanaslou,M., Aguayo,P., Armstrong,D., Basit,M.,
Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
Grant,O., Hanner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,
Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
HTGS Submission
Unpublished
2 (bases 1 to 116917)
Evans,G.A., Athanaslou,M., Bradbury,P., Brignac,S., Bumeister,R.,
Davis,C., English,C., Franklin,T.L., Garner,H.R., Gordon,M.,
Gotway,G., Grant,O., Hanner,L., Harris,J., Hinson,S.,
Narayanaswamy,U., Newton,J., O'Brien,K., Oliver,T., Patel,P.,
Probst,S., Rayner,S., Schageman,J., Schilling,P., Schultz,R.,
Syed,M., Valenzuela,D., Ward,T. and Wilson,R.
Direct Submission
Submitted (21-OCT-1997) Genome Science and Technology Center,
University of Texas Southwestern Medical Center at Dallas, 5323
Harry Hines Blvd, Dallas, TX 75235-8591, USA
3 (bases 1 to 116917)
Evans,G.A., Athanaslou,M., Basit,M., Bradbury,P., Brignac,S.,
Bumeister,R., Davis,C., English,C., Franklin,T.L., Garner,H.R.,
Gee,V., Gordon,M., Gotway,G., Grant,O., Hanner,L., Harris,J.,
Hinson,S., Narayanaswamy,U., Newton,J., O'Brien,K., Patel,P.,
Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,

```

```

TITLE
JOURNAL
Direct Submission
Submitted (17-DEC-1997) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd, Dallas, TX 75235-8591, USA
REFERENCE
4 (bases 1 to 116917)
AUTHORS
Evans,G.A., Athanaslou,M., Aguayo,P., Armstrong,D., Basit,M.,
Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
Grant,O., Hanner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,
Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
TITLE
JOURNAL
Submitted (16-SEP-1998) Genome Science & Technology Center,
University of Texas Southwestern Medical Center, 5323 Harry Hines
Blvd, Dallas, TX 75235-8591, USA
COMMENT
On Sep 16, 1998 this sequence version replaced gi:26955560.
Further information regarding the map of this region or
annotation of pDJ41616 can be found at
http://gestec.swmed.edu/chromos05.htm.
IMPORTANT: This submission contains the entire insert of clone
pDJ41616. pDJ41616 comes from the RPI-3 PAC library constructed
at the Fowell Park Cancer Institute by the Pieter de Jong group.
This clone has been finished according to strict quality criteria
and attempts have been made to resolve all base calling problems
such as compressions and repetitive elements. The expected
Phred/Phrap calculated errors/10kb is 0.56. In addition, attempts
have been made to assure over 99% of consensus base calls consist
of either double-stranded coverage or 2 types of labeling chemistry
on one strand.
CHROMOSOMAL LOCUS: This PAC clone comes from the Chromosome 15q26.1
Bloom's syndrome/DNA polymerase gamma region. This region is
mapped between STSS D15S171 and D15S652
MARKER CONFIRMATION: STSS sequence confirmed; WI-6813, D15S963,
SHGC-9392.
MAPPED CLONE OVERLAP: pDJ422e7.
FEATURES
Source
1..116917
/organism="Homo sapiens"
/db_xref="taxon:9606"
repeat_region 5376..5521 /rpt_family="MIR"
repeat_region complement(8545..8834)
repeat_region 9913..10219 /rpt_family="Alu"
repeat_region 11974..12151 /rpt_family="MER2"
repeat_region 13289..16374 /rpt_family="MLT1"
repeat_region 17666..17816 /rpt_family="L1"
repeat_region 18471..18642 /rpt_family="L1"
repeat_region complement(20327..20615)
repeat_region complement(20947..21077)
repeat_region 21210..21508 /rpt_family="MIR"
repeat_region 23183..23483 /rpt_family="Alu"
repeat_region complement(26948..27241)
repeat_region complement(27278..27386)
repeat_region 30624..31046 /rpt_family="MLT1"
repeat_region complement(32058..32119)
repeat_region /rpt_family="MER43"
repeat_region complement(32746..32980)
repeat_region /rpt_family="MER43"
repeat_region 33389..37165

```

```
repeat_region /rpt_family="L1" 37567.38347
repeat_region /rpt_family="MER42" complement(43991.44279)
repeat_region /rpt_family="Alu" complement(44423.44623)
repeat_region /rpt_family="MIR" complement(47160.47446)
repeat_region /rpt_family="Alu" complement(49643.49937)
repeat_region /rpt_family="Alu" 50664.50947
repeat_region /rpt_family="Alu" complement(52207.52266)
repeat_region /rpt_family="MER46" complement(52470.52541)
repeat_region /rpt_family="MER46" complement(58313.58613)
repeat_region /rpt_family="Alu" 59242.59382
repeat_region /rpt_family="MER46" complement(59243.59474)
repeat_region /rpt_family="MER46" 61236.61507
repeat_region /rpt_family="Alu" 63584.63715
repeat_region /rpt_family="MERS" complement(63584.63714)
repeat_region /rpt_family="MERS" complement(70158.70391)
repeat_region /rpt_family="Alu" 72078.72148
repeat_region /rpt_family="L1" 72205.72466
repeat_region /rpt_family="Alu" 74340.74493
repeat_region /rpt_family="MIR" complement(76377.76469)
repeat_region /rpt_family="MIR" 77426.77717
repeat_region /rpt_family="Alu" complement(82461.82745)
repeat_region /rpt_family="Alu" 84074.84249
repeat_region /rpt_family="Alu" complement(85332.85612)
repeat_region /rpt_family="Alu" 88044.88333
repeat_region /rpt_family="Alu" 90721.90757
repeat_region /rpt_family="MIR" 93074.93360
repeat_region /rpt_family="Alu" complement(94986.95294)
repeat_region /rpt_family="Alu" 98459.98706
repeat_region /rpt_family="L1" complement(98873.99025)
repeat_region /rpt_family="MIR" complement(99895.100145)
repeat_region /rpt_family="Alu" 106174.100500
repeat_region /rpt_family="THL1" complement(100506.100539)
repeat_region /rpt_family="Alu" 101454.101641
repeat_region /rpt_family="MLT1" complement(102058.102275)
repeat_region /rpt_family="Alu" 106631.106929
repeat_region /rpt_family="Alu" complement(109874.111578)
repeat_region /rpt_family="L1"

repeat_region complement(112332.112482)
repeat_region /rpt_family="MERS" complement(112818.113102)
repeat_region /rpt_family="Alu"

BASE COUNT 36124 a 24604 c 23503 g 32686 t
ORIGIN

Query Match 2.0%; Score 21; DB 9; Length 116917;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 786 tatttggagatgaagctcttta 806
|||||
Db 27620 TATTGGAGATGAAGTCTTTA 27600

RESULT 9
AC002421 131047 bp DNA HTG 12-AUG-1997
LOCUS Homo sapiens chromosome X clone pWXL1, *** SEQUENCING IN PROGRESS
DEFINITION ***; 4 unordered pieces.
ACCESSION AC002421
VERSION AC002421.1 GI:2323248
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 131047)
AUTHORS Chen, E., Brownstein, B.H., States, D.J., Schlessinger, D. and
Mazzarella, R.
TITLE Direct Submission
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 131047)
AUTHORS Brownstein, B.H., States, D.J. and Mazzarella, R.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,
Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA
COMMENT Current status of this project is available at:
'http://genome.wustl.edu/cgm/seq_projects.html'
Submitted by:
Elliott Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp.,
850 Lincoln Center Drive,
Foster City, CA 9404 USA
e-mail: ellison@genseq.apldbio.com
and
Buddy Brownstein,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232
4566 Scott Avenue,
St. Louis, MO 63110, USA
e-mail: buddy@genetics.wustl.edu
and
David J. States,
Institute for Biomedical Computing
Washington University in St. Louis
700 South Euclid Ave
St. Louis, MO 63108 USA
e-mail: states@bc.wustl.edu.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```

\* as soon as it is available and the accession number will  
 \* be preserved.

1 22595: contig of 22595 bp in length  
 \* 22596 66883: contig of 44288 bp in length  
 \* 66884 122647: contig of 55764 bp in length  
 \* 122648 131047: contig of 8400 bp in length.

# FEATURES

Location/Qualifiers  
 1. .131047  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="pWXD1"  
 /chromosome="X"

BASE COUNT 40590 a 23040 c 23923 g 43494 t  
 ORIGIN

Query Match 2.0% Score 21; DB 27; Length 131047;

Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 907 cttgtgtgtgtgtgagagaca 927

Db 33367 CTTGTGTGTGTGAGAGACA 33387

# RESULT 10

AC023308 136410 bp DNA HTG 07-JUN-2000  
 LOCUS Homo sapiens chromosome 17 clone RP11-160A21 map 17, WORKING DRAFT  
 DEFINITION SEQUENCE, 12 unordered pieces.  
 ACCESSION AC023308  
 VERSION AC023308.3 GI:8318560  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 136410)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 17, clone RP11-160A21

Unpublished

2 (bases 1 to 136410)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavskiy,L.,  
 Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
 Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,  
 Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., Lander,E., Lander,T., Largocque,K., Lehoczy,J., Levine,R.,  
 Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,  
 McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,  
 Menues,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M.,  
 Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,  
 Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,  
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,  
 Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,  
 Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and  
 Zody,M.

Direct Submission

Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 7, 2000 this sequence version replaced gi:7652825.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L5415

Center clone name: 160\_A\_21

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 130690 bases at least Q40  
 Consensus quality: 133949 bases at least Q30  
 Consensus quality: 134881 bases at least Q20  
 Insert size: 145000; agarose-fp  
 Insert size: 135310; sum-of-contigs  
 Quality coverage: 4.3 in Q20 bases; agarose-fp  
 Quality coverage: 4.6 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 12 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1. 1112: contig of 1112 bp in length  
 1113 1212: gap of 100 bp  
 1213 3802: contig of 2590 bp in length  
 3803 3902: gap of 100 bp  
 3903 9015: contig of 5113 bp in length  
 9016 9115: gap of 100 bp  
 9116 17740: contig of 8625 bp in length  
 17741 17840: gap of 100 bp  
 17841 27210: contig of 9370 bp in length  
 27211 27310: gap of 100 bp  
 27311 35885: contig of 8275 bp in length  
 35886 35685: gap of 100 bp  
 35686 43357: contig of 7672 bp in length  
 43358 43457: gap of 100 bp  
 43458 52798: contig of 9341 bp in length  
 52799 52898: gap of 100 bp  
 52899 72838: contig of 19940 bp in length  
 72839 72938: gap of 100 bp  
 72939 90225: contig of 17287 bp in length  
 90226 90325: gap of 100 bp  
 90326 110518: contig of 20193 bp in length  
 110519 110618: gap of 100 bp  
 110619 136410: contig of 25792 bp in length.

# FEATURES

## SOURCE

1. .136410  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="RP11-160A21"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1. .1112  
 /note="assembly\_fragment"  
 1213. 3802  
 /note="assembly\_fragment"  
 3903. 9015  
 /note="assembly\_fragment"  
 9116. 17740  
 /note="assembly\_fragment"  
 17841..27210  
 /note="assembly\_fragment"  
 27311..35585  
 /note="assembly\_fragment"  
 35686..43357  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:right  
 43458..52798

## misc\_feature



```

repeat_region      complement(24242..24515)
                    /rpt_family="MIR"
repeat_region      complement(24773..25066)
                    /rpt_family="AluSg"
repeat_region      complement(25862..26027)
                    /rpt_family="MLT1C"
repeat_region      complement(26021..26165)

Query Match       2.0% Score 21; DB 10; Length 152355;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 786 tatitggagatgaagtcttta 806
    |||||
Db 1356 TATTGGAGATGAAGTCTTTA 1376

RESULT 12
AC018652/c
LOCUS          AC018652.3 167873 bp DNA
DEFINITION    Homo sapiens clone RP11-28022, WORKING DRAFT SEQUENCE, 10 unordered
               pieces.
ACCESSION     AC018652
VERSION       AC018652.3 GI:7382211
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 167873)
AUTHORS       Birren,E., Linton,L., Nusbaum,C. and Lander,E.
TITLE         Homo sapiens, clone RP11-28022
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 167873)
AUTHORS       Birren,E., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
               Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
               Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
               Choepel,X., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
               DeArelle,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
               Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
               Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
               Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
               Landers,T.C., Lechoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
               Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
               Meldrim,J., Menes,L., Morrow,J., Naylor,J., Norman,C.H.,
               O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K., Pierre,N.,
               Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,
               Santos,F., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
               Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
               Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A.
               and Zody,M.
               Direct Submission
               Submitted (15-DEC-1999) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141 USA
               On Apr 1, 2000 this sequence version replaced gi:6649326.
               All repeats were identified using RepeatMasker:
               http://ftp.genome.washington.edu/RM/RepeatMasker.html
               ----- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ----- Project Information
               Center project name: L4802
               Center clone name: 28_O_22
               ----- Summary Statistics
               Sequencing vector: M13; M7815; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 162467 bases at least Q40
               Consensus quality: 165076 bases at least Q30
               Consensus quality: 165968 bases at least Q20

TITLE
JOURNAL
COMMENT

```



Mon Nov 6 10:14:09 2000

```

* 42573 47699: contig of 5127 bp in length
* 47700 54016: contig of 6317 bp in length
* 54017 60547: contig of 6531 bp in length
* 60548 68280: contig of 7733 bp in length
* 68281 75214: contig of 6934 bp in length
* 75215 80915: contig of 5701 bp in length
* 80916 87222: contig of 6307 bp in length
* 87223 94562: contig of 7340 bp in length
* 94563 106577: contig of 12015 bp in length
* 106578 119209: contig of 12632 bp in length
* 119210 141857: contig of 22648 bp in length
* 141858 172091: contig of 30234 bp in length.
FEATURES
    source
        1. .172091
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="2511.J.5"
        /clone_lib="CIB Human BAC"
        /map="17"
        /chromosome="17"
BASE COUNT 44773 a 39840 c 42000 g 44968 t 510 others
ORIGIN

```

```

Query Match 2.08; Score 21; DB 27; Length 172091;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 786 tatttgagatgaagctcttta 806
|||||
DB 21668 TATTGGAGATGAGCTCTTA 21668

```

```

RESULT 14
LOCUS HSP0071 3907 bp mRNA PRI 18-APR-1997
DEFINITION H. sapiens mRNA for p0071 protein.
ACCESSION X81889
VERSION X81889.1 GI:1702923
KEYWORDS armadillo protein; p0071 protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 3907)
Hatzfeld, M. and Nachtsheim, C.
Cloning and characterization of a new armadillo family member,
p0071, associated with the junctional plaque: evidence for a
subfamily of closely related proteins
J. Cell. Sci. 109 (Pt 11), 2767-2778 (1996)
JOURNAL 97092329
MEDLINE 2 (bases 1 to 3907)
REFERENCE Hatzfeld, M.
Direct Submission
AUTHORS Submitted (13-AUG-1996) M. Hatzfeld, Max Planck Inst. for
JOURNAL Biophysical Chem., PO Box 2841, D-37018 Goettingen, FRG
COMMENT Related sequences:
EST UXT00071, and M62015.
FEATURES
    source
        1. .3907
        /organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/sex="female"
/dev_stage="85 years old"
/tissue_type="brain, frontal cortex"
/clone_lib="lambda ZAP II (Stratagene)"
142..3777
/notes="member of armadillo multigene family; component of
the junctional plaque of desmosomes"
/codon_start=1
/product="p0071 protein"
/protein_id="CAA57478.1"
/db_xref="GI:1702924"
/db_xref="SPTREBL:Q95959"
/translation="MPAEQASLVVEEGQPGQPTQROEAASTGFMETATTIILASKEQ
ELQQLTRLEVERQIVASOLERCLGAEPSIASTSTSEKFSFWRSDVDPNTGYSK
PRVDVQPNYLIETPEQGLYSPQSLHSEGLSGNSRSTOMNSYSDSGYQEA
GSFHSNQVRADNRQHSFISGSGFSGSVDPRLNPNFSAISSTILPAARASPS
AQSPSYIVISGVSFSGSLRTSLGSGFSGSVDPRLNPNFSAISSTILPAARASPS
QRPASPTAIRIGSVTSQTSNPNPTQYQTTARVGSPLTLDQATRVASQGOVG
SSPKRSQMTAVPQHLGLPQLRTVHDMQEQGQOQDIYERMPVPPRPSLUTGLRSYAS
QHSQGLDLSAVSPDLHITPIYEGRTYSPVYRPNHGTVEQSGQALYRTGVSIGI
GNLQTSQSRTLVQRYNVALNTATVAPYRPIQYRVOECYNNRLOHAPVADDTT
RSPSIDSKQDPREFARDPELPEVILHMEHQEPFVQVAAVLOHLCFGDNKVMKEV
CRLGIEKHLVDLDRVLEVNACGALNLYFGKSTDENKIAMKNVGGIGPALLRLR
KSIDAEVRELVTVLWNLSSCDVAKMTIIRDALSTLTNTVIVPHSGHNSFDDDKHI
KFQTSVLRLNTTGLRLNLTSAEAKOMRSCGLVDSLLIVHTCVNTSDISKTYVE
NCVCTLRNLVLELVQARLLGLNELDLGLKESPKDSEPCSKKKKKKRTTPQ
EDWDGVPGLGSKPKGKGLPIIVELLRMDNDRVSSGATARNMALDVNKKELIKYAM
NWKFAAYIRGGPKPKGLPIIVELLRMDNDRVSSGATARNMALDVNKKELIKYAM
RDLVARKLPGGSPSVLSETHAACCALHEVTSKMENAKALADSGGIEKLIVNITGR
GDRSLKVVYKAAQVNLWQIRLRSYIKKDKGNQNHFIIPVSTLERDFKSHPSLS
TTNQMSPIIGSVGSSPALLGIRDPRSDYDRTQPMQVITVPSQGGATHKGLYPGSS
KPSPIYISYSSPAREQNRRLHQQLYISODSNRKNFDRYRLYLQSPHSTEDYDFD
RVHFPASTDYSTQYGLKSTNNVDFYKTPRSYAEQYPGSPDSWVYDQDAQRNSFF
LTLRLR"
BASE COUNT 1083 a 1001 c 955 g 868 t
ORIGIN

```

```

Query Match 1.9%; Score 20; DB 82; Length 3907;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 756 aaggaagaaacacacactt 775
|||||
DB 3856 AAGGAAAAAACAACACTT 3837

```

```

RESULT 15
AC019948 38182 bp DNA HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC019948
VERSION AC019948.1 GI:6664949
KEYWORDS HTG; HTGS-PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 38182)
Adams, M. and Venter, J. C.
Direct Submission
TITLE Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
JOURNAL Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211494 by the submitter.
For more information on this record e-mail to fty@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
    Location/Qualifiers

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:35:23 ; Search time 255.06 Seconds  
(without alignments)  
1558.265 Million cell updates/sec

Title: US-09-101-423A-2  
Perfect score: 1058  
Sequence: 1 attctgtgagcctattagc.....acacattcgagcgttggtc 1058

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 480022 seqs, 187831343 residues

Word size : 0

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1058	100.0	1058	18 T72786	Metastasis inducin
2	20	1.9	3907	18 T79968	Presenilin-interac
3	20	1.9	3907	21 A09308	Human p007i coding
4	19	1.8	213	20 X41398	Human secreted pro
5	19	1.8	1150	21 X57077	Human secreted pro
6	19	1.8	1193	21 X97127	Human secreted pro
7	19	1.8	1269	20 X80674	Clone nt746.4 enco
8	18	1.7	322	21 A31348	plant microsatelli
9	18	1.7	433	21 A31378	Sequence of the gl
10	18	1.7	1325	4 N30074	Transcription term
11	18	1.7	1325	5 N40213	Enterococcus faeca
12	18	1.7	1657	20 X13303	

c 13	18	1.7	2178	4 N30016	Sequence of the pr
c 14	18	1.7	2472	18 T74953	Staphylococcus aur
c 15	18	1.7	3505	18 T74667	pig myogenin gene.
c 16	18	1.7	4091	11 Q04035	stem cell leukaemi
c 17	18	1.7	4199	11 Q05330	Peripheral nervous
c 18	18	1.7	6371	17 T30194	Human WRN genomic
c 19	18	1.7	6404	17 T30195	Compound simple se
c 20	18	1.7	87350	18 X83003	Repeat sequence fr
c 21	17	1.6	18	17 T30414	Microsatellite seq
c 22	17	1.6	33	18 T65770	Repeat sequence fr
c 23	17	1.6	40	13 Q33618	Microsatellite seq
c 24	17	1.6	50	18 T65791	Repeat sequence fr
c 25	17	1.6	56	13 Q33603	Sequence of a micr
c 26	17	1.6	56	18 T65707	Repeat sequence fr
c 27	17	1.6	58	13 Q34116	Repeat sequence fr
c 28	17	1.6	60	18 T65762	Microsatellite seq
c 29	17	1.6	62	13 Q33779	Microsatellite seq
c 30	17	1.6	63	13 Q33594	Microsatellite seq
c 31	17	1.6	64	13 Q33534	DNA polymerase lig
c 32	17	1.6	78	18 T59165	DNA polymerase lig
c 33	17	1.6	78	18 T59167	DNA polymerase lig
c 34	17	1.6	78	18 T59168	Microsatellite seq
c 35	17	1.6	86	13 Q34050	Repeat sequence fr
c 36	17	1.6	91	18 T65740	(dc-da)n.(dg-dt)n
c 37	17	1.6	92	18 T66081	Microsatellite nuc
c 38	17	1.6	101	21 A35574	Microsatellite nuc
c 39	17	1.6	123	21 A35573	Microsatellite nuc
c 40	17	1.6	143	21 A35575	2-chromosomal micr
c 41	17	1.6	198	19 V48138	Polymorphic repeat
c 42	17	1.6	210	18 T65674	Polymorphic repeat
c 43	17	1.6	228	18 T65694	Plant microsatelli
c 44	17	1.6	251	21 A31344	Myrtaceae microsat
c 45	17	1.6	260	21 A35343	

ALIGNMENTS

RESULT 1  
T72786  
ID T72786 standard; DNA; 1058 BP.  
XX  
AC T72786;  
XX  
DT 22-SEP-1997 (first entry)  
XX  
DE Metastasis inducing DNA C5.  
XX  
KW Metastasis-inducing DNA; Met-DNA; cancer; diagnosis;  
KW Osteopontin; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9725443-A1.  
XX  
PD 17-JUL-1997.  
XX  
PF 10-JAN-1997; 97WO-GB00074.  
XX  
PR 10-JAN-1996; 96GB-0000470.  
XX  
PA (UYLI-) UNIV LIVERPOOL.  
XX  
PI Barracough BR, Rudland PS;  
XX  
DR WPI; 1997-372878/34.  
XX  
PT New isolated metastasis-inducing DNA - used to develop products to  
PT identify and treat patients at risk from metastatic tumours  
XX  
PS Claim 9; Page 25; 38pp; English.  
XX  
CC Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12

CC and C20 (T72785-90) are entirely novel short stretches of human  
 CC regulatory DNA capable of inducing metastasis. They were  
 CC identified using a method for detecting Met-DNA that involves  
 CC transferring human DNA from a malignant, metastatic cancer cell  
 CC (in this case, breast cancer) into a cell line (pref. rat Rama 37)  
 CC that produces only benign, non-metastasizing tumours when injected  
 CC into a syngeneic animal, selecting those animals having metastasizing  
 CC syngeneic animal, selecting those animals having metastasizing  
 CC tumours, and recovering the Met-DNA from them. The isolated  
 CC Met-DNAs can be used to develop products to identify and treat  
 CC patients at risk from metastatic tumours.  
 XX  
 SQ Sequence 1058 BP; 320 A; 219 C; 249 G; 270 T; 0 other;

Query Match 100.0%; Score 1058; DB 18; Length 1058;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1058; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgctgtgagcctattagcgaatttggtagcgcctcttttaagggtgtagatacaag 60  
 DB 1 attgctgtgagcctattagcgaatttggtagcgcctcttttaagggtgtagatacaag 60  
 QY 61 aatgggttgaattctgtcccaaacgctctccatgttttccaaattacacttgcacac 120  
 DB 61 aatgggttgaattctgtcccaaacgctctccatgttttccaaattacacttgcacac 120  
 QY 121 tgtgttcagcagccagaatttggtagtggtagcaggggtcggtgggaagagagag 180  
 DB 121 tgtgttcagcagccagaatttggtagtggtagcaggggtcggtgggaagagagag 180  
 QY 181 ggtaaagaaacacagcagcttaagtccaaacagctccagagagactatctgtagaata 240  
 DB 181 ggtaaagaaacacagcagcttaagtccaaacagctccagagagactatctgtagaata 240  
 QY 241 acatcagacattgagagaaattgtagtattgttttcaatgggtatcccaaggaact 300  
 DB 241 acatcagacattgagagaaattgtagtattgttttcaatgggtatcccaaggaact 300  
 QY 301 ttcacatcgtataaaataaattactgctggcactaaatcccaattggaaatgccccacaca 360  
 DB 301 ttcacatcgtataaaataaattactgctggcactaaatcccaattggaaatgccccacaca 360  
 QY 361 atttatctccacttcagctgctaccatagctgctgagtggtgagagagagagagagag 420  
 DB 361 atttatctccacttcagctgctaccatagctgctgagtggtgagagagagagagagag 420  
 QY 421 tccgctctgataaaatagactttgttaataatttggagacgggagctctgtgacagga 480  
 DB 421 tccgctctgataaaatagactttgttaataatttggagacgggagctctgtgacagga 480  
 QY 481 acagtcacaaacggcctgttattcatgttcccgatagagagagagagagagagagag 540  
 DB 481 acagtcacaaacggcctgttattcatgttcccgatagagagagagagagagagagag 540  
 QY 541 ccccaaaacagtcaggtgctgtaatttccctccatgaagccttcttcacaaattagcaa 600  
 DB 541 ccccaaaacagtcaggtgctgtaatttccctccatgaagccttcttcacaaattagcaa 600  
 QY 601 ccattgaggaagcagctgctgactgttaccacaaagtggaactttccaaagagacacat 660  
 DB 601 ccattgaggaagcagctgctgactgttaccacaaagtggaactttccaaagagacacat 660  
 QY 661 atattgagcaagacatttctggtgctgctgctgctgctgctgctgctgctgctgctgct 720  
 DB 661 atattgagcaagacatttctggtgctgctgctgctgctgctgctgctgctgctgctgct 720  
 QY 721 ttattaaactgcttcttcttgaacacccactcaagaaaaaaacacacacttaggt 780  
 DB 721 ttattaaactgcttcttcttgaacacccactcaagaaaaaaacacacacttaggt 780  
 QY 781 gacattatttgagagaaagctctttatagatgcttaatttaacagagactttaag 840  
 DB 781 gacattatttgagagaaagctctttatagatgcttaatttaacagagactttaag 840

DB 781 gacattatttgagagaaagctctttatagatgcttaatttaacagagactttaag 840  
 QY 841 ccggtctctatccatttaataatgaatggtgtccctacaagaagaactgggacagagta 900  
 DB 841 ccggtctctatccatttaataatgaatggtgtccctacaagaagaactgggacagagta 900  
 QY 901 tgtacacttg-gtgtgtgtgagagacaacgtgagagagctgaagagagacgtacaaagt 960  
 DB 901 tgtacacttg-gtgtgtgtgagagacaacgtgagagagctgaagagagacgtacaaagt 960  
 QY 961 agagaaagggc-gacccttattcacactgagacaacacagctcatgtgtggtcgatagta 1020  
 DB 961 agagaaagggc-gacccttattcacactgagacaacacagctcatgtgtggtcgatagta 1020  
 QY 1021 gagtatcccccaagactcacacatttcgaacgttggtc 1058  
 DB 1021 gagtatcccccaagactcacacatttcgaacgttggtc 1058

## RESULT 2

T79968/c  
 ID T79968 standard: cDNA; 3907 BP.

XX T79968;

XX 06-FEB-1998 (first entry)

XX Presenilin-interacting protein p0071 cDNA.

XX Presenilin-interacting protein; human; Alzheimer's disease;

XX diagnosis; therapy; transgenic animal; animal model; p0071; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 142..3777

XX /\*tag= a

XX W09727296-Al.

XX 31-JUL-1997.

XX 27-JAN-1997; 97WO-CA00051.

XX 02-JAN-1997; 97US-0034590.

XX 26-JAN-1996; 96US-0592541.

XX 05-JUL-1996; 96US-0021673.

XX 12-JUL-1996; 96US-0021700.

XX 08-NOV-1996; 96US-0029895.

XX (HSCR-) HSC RES & DEV LP.

XX (UTOR) UNIV TORONTO GOVERNING COUNCIL.

XX Fraser PE, Rommens JM, St George-Hyslop PH;

XX WPI: 1997-393684/36.

XX P-PSDB; W24560.

XX Presenilin-interacting protein genes - used to develop products for

XX the diagnosis, therapy and study of Alzheimer's disease and related

XX disorders

XX Claim 2; Page 104-107; 133pp; English.

XX This nucleic acid sequence includes a coding sequence for p0071

XX (see W24560), a human protein with 'armadillo' repeats, identified

XX as a presenilin-interacting protein (PIP). A yeast two-hybrid kit

XX was employed to screen a human adult brain cDNA library for clones

XX which interacted with presenilin transmembrane 6-7 loop domain;

XX mutations in this loop domain are known to be causative of

XX Alzheimer's disease (AD). 9 PIP gene sequences (see T79966-74)

XX including p0071 were identified. Each represents a new therapeutic

XX target for the treatment of AD. PIP nucleic acids, PIP proteins



SQ Sequence 213 BP; 62 A; 37 C; 30 G; 83 T; 1 other;

Query Match 1.8%; Score 19; DB 20; Length 213;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 760 aaaaaaacacacttagg 778  
|||||  
Db 195 aaaaaaacacacttagg 213

## RESULT 5

297077/c  
ID 297077 standard; cDNA; 1150 BP.

AC 297077;

DT 19-APR-2000 (first entry)

DE Human secreted protein gene 59 cDNA clone HDPXR23, SEQ ID NO:69.

XX Human; secreted protein; cancer; tumour; developmental abnormality;  
KW foetal deficiency; blood disorder; immune system disorder; inflammation;  
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;  
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;  
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;  
KW digestive disorder; endocrine disorder; infection; AIDS; leukaemia;  
KW therapy; chromosome 9; ds.

XX Homo sapiens.

XX WO9966041-A1.

XX 23-DEC-1999.

XX 15-JUN-1999; 99WO-US13418.

XX 16-JUN-1998; 98US-0089507.

XX 16-JUN-1998; 98US-0089508.

XX 16-JUN-1998; 98US-0089509.

XX 16-JUN-1998; 98US-0089510.

XX 22-JUN-1998; 98US-0090112.

XX 22-JUN-1998; 98US-0090113.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Rosen CH, Wei Y, Young PE, Florence KA;

PI Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R;

PI Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;

XX WPI; 2000-106100/09.

XX P-PSDB; Y86273.

XX New isolated human genes and the secreted polypeptides they encode,  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders -

PS Claim 1; Page 352; 586pp; English.

XX 297019 to 297137 represent 94 isolated human secreted protein genes.  
CC Y86215 to Y86333 are the secreted proteins encoded by the 94 human genes.  
CC This sequence represents a fragment of one of the human secreted  
CC proteins. The genes and their corresponding secreted polypeptides are  
CC useful for preventing, treating or ameliorating medical conditions,  
CC e.g., by protein or gene therapy. Also pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new genes. Specific  
CC uses are described for each of the 94 genes, based on which tissues they  
CC are most highly expressed in, and include developing products for the  
CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
CC and foetal deficiencies, blood disorders, diseases of the immune system,  
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive

CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin  
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
CC disorders, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
CC The sequences shown in Y86334 to Y86585 represent fragments of the  
CC secreted proteins.

XX SQ Sequence 1150 BP; 407 A; 196 C; 244 G; 303 T; 0 other;

## Query Match

Best Local Similarity 1.8%; Score 19; DB 21; Length 1150;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 ggaagaagaaggaagga 184

|||||  
Db 610 GGAAGAAGGAGGAGGTA 592

## RESULT 6

297127/c

ID 297127 standard; cDNA; 1193 BP.

XX AC 297127;

XX 19-APR-2000 (first entry)

XX Human secreted protein gene 59 cDNA clone HDPXR23, SEQ ID NO:119.

XX Human; secreted protein; cancer; tumour; developmental abnormality;

XX foetal deficiency; blood disorder; immune system disorder; inflammation;

XX autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;

XX schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;

XX atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;

XX digestive disorder; endocrine disorder; infection; AIDS; leukaemia;

XX therapy; chromosome 9; ds.

XX Homo sapiens.

XX WO9966041-A1.

XX 23-DEC-1999.

XX 15-JUN-1999; 99WO-US13418.

XX 16-JUN-1998; 98US-0089507.

XX 16-JUN-1998; 98US-0089508.

XX 16-JUN-1998; 98US-0089509.

XX 16-JUN-1998; 98US-0089510.

XX 22-JUN-1998; 98US-0090112.

XX 22-JUN-1998; 98US-0090113.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Rosen CA, Wei Y, Young PE, Florence KA;

PI Soppet DR, Brewer LA, Endress GA, Carter KC, Mucenski M, Ebner R;

PI Lafleur DW, Olsen HS, Shi Y, Moore PA, Komatsoulis G;

XX WPI; 2000-106100/09.

XX P-PSDB; Y86323.

XX New isolated human genes and the secreted polypeptides they encode,  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders -

PS Claim 1; Page 382; 586pp; English.

XX 297019 to 297137 represent 94 isolated human secreted protein genes.  
CC Y86215 to Y86333 are the secreted proteins encoded by the 94 human genes.  
CC This sequence represents a fragment of one of the human secreted  
CC proteins. The genes and their corresponding secreted polypeptides are  
CC useful for preventing, treating or ameliorating medical conditions, be  
CC e.g., by protein or gene therapy. Also pathological conditions can be

Mon Nov 6 10:14:10 2000

CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new genes. Specific  
 CC uses are described for each of the 94 genes, based on which tissues they  
 CC are most highly expressed in, and include developing products for the  
 CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
 CC and foetal deficiencies, blood disorders, diseases of the immune system,  
 CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive  
 CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin  
 CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
 CC disorders, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The sequences shown in Y86334 to Y86585 represent fragments of the  
 CC secreted proteins.  
 XX  
 XX Sequence 1193 BP; 402 A; 216 C; 255 G; 318 T; 2 other;  
 SQ  
 Query Match 1.8%; Score 19; DB 21; Length 1193;  
 Best Local Similarity 100.0%; Pred. No. 4.6;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 166 ggaagaagaagaaggta 184  
 |||||  
 Db 656 ggaagaagaagaaggta 638  
 |||||  
 RESULT 7  
 ID X80674/c  
 XX X80674 standard; cDNA; 1269 BP.  
 AC X80674;  
 XX  
 DT 22-OCT-1999 (first entry)  
 XX  
 DE Clone nt746\_4 encoding secreted protein-nt746\_4.  
 XX  
 DE Secreted protein; cytokine; cell proliferation; immune stimulation;  
 XX vaccine; immune suppression; haematopoiesis; tissue growth; activin;  
 KW inhibitor; chemotaxis; chemokinesis; haemostasis; thrombolytic; ss;  
 KW receptor; ligand; anti-inflammatory; cadherin; tumour; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH 133..435  
 FT CDS /tag= a  
 FT /product= "Secreted protein nt746\_4"  
 FT  
 XX WO928335-A1.  
 XX 10-JUN-1999.  
 XX  
 XX 02-DEC-1998; 98WO-US25512.  
 XX  
 XX 30-NOV-1998; 98US-0203106.  
 XX 04-DEC-1997; 97US-0067454.  
 XX  
 XX (GEM) GENETICS INST INC.  
 XX  
 XX Agostino MJ, Clark HF, Collins-Racie LA, Evans C;  
 PI Pechtel K, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 PI Steininger RJ, Treacy M, Wong G;  
 XX  
 XX WPI; 1999-385352/32.  
 DR P-PSDB; Y26039.  
 XX  
 XX New polynucleotides encoding secreted human proteins  
 PT  
 XX Claim 34; Pages 114-115; 124pp; English.  
 PS  
 XX The present sequence is a known clone nt746\_4 (deposited as ATCC  
 CC 98600) isolated from human adult brain cDNA library. It encodes  
 CC secreted protein nt746\_4. Recombinant secreted proteins can be produced

CC by transforming host cells and culturing them under suitable conditions.  
 CC The polynucleotide and protein are predicted to have biological  
 CC activities which would make them suitable for treating, preventing or  
 CC ameliorating medical conditions in humans and animals. Some predicted  
 CC biological activities include cytokine and cell proliferation/  
 CC differentiation activity, immune stimulating (e.g. as vaccines) or  
 CC suppressing activity, haematopoiesis regulating activity, tissue growth  
 CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,  
 CC hemostatic and thrombolytic activity, receptor/ligand activity, anti-  
 CC inflammatory activity, cadherin/tumour invasion suppressor activity, and  
 CC tumour inhibition activity. The polynucleotide can be used for gene  
 XX therapy.  
 XX Sequence 1269 BP; 421 A; 237 C; 292 G; 319 T; 0 other;  
 SQ  
 Query Match 1.8%; Score 19; DB 20; Length 1269;  
 Best Local Similarity 100.0%; Pred. No. 4.6;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 166 ggaagaagaagaaggta 184  
 |||||  
 Db 724 ggaagaagaagaaggta 706  
 |||||  
 RESULT 8  
 ID A31348 standard; DNA; 322 BP.  
 XX A31348;  
 AC A31348;  
 XX  
 DT 05-JUL-2000 (first entry)  
 XX  
 DE Plant microsatellite marker #309.  
 XX  
 XX Plant microsatellite sequence; core repeat sequence; detection; probe;  
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;  
 KW variety identification; genetic variability evaluation; primer; ss.  
 XX  
 OS Eucalyptus grandis.  
 XX  
 XX WO9967421-A1.  
 XX  
 XX 29-DEC-1999.  
 XX  
 XX 25-JUN-1999; 99WO-NZ00092.  
 XX  
 XX 25-JUN-1998; 98US-0105307.  
 XX  
 XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.  
 XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX  
 XX Havukkala IJ, Bloksberg LN, Glenn M;  
 XX  
 XX WPI; 2000-116958/10.  
 XX  
 XX New plant microsatellite markers and associated flanking species for  
 PT the detection of polymorphic genetic markers -  
 PS  
 XX Claim 1; Page 168; 392pp; English.  
 XX  
 XX Sequences A31040-A32093 represent novel plant microsatellite sequences  
 CC and associated flanking species. The sequences comprise a central core  
 CC repeat sequence, especially selected from the sequences A32094-A32096  
 CC with left and right flanking sequences. The polynucleotide sequences  
 CC can be used in the detection of DNA polymorphisms, in genome mapping,  
 CC in physical mapping, in positional cloning of genes, in variety  
 CC identification and in evaluation of genetic variability within and  
 CC between plant tissues, populations, cultivars, species and species  
 CC groups. They may also be used to design hybridization probes for  
 CC oligonucleotide fingerprinting and library screening and to design  
 CC primers for microsatellite-primed PCR. Microsatellite markers are  
 CC useful to locate specific economically useful genes in plant genomes.

XX SQ Sequence 322 BP; 70 A; 87 C; 91 G; 74 T; 0 other;

Query Match  
Best Local Similarity 1.7%; Score 18; DB 21; Length 322;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 908 ttgtgtgtgtgtgagaga 925  
|||||

Db 4 ttgtgtgtgtgtgagaga 21

RESULT 9

A31378 ID A31378 standard; DNA; 433 BP.

XX AC A31378;

DE 05-JUL-2000 (first entry)

XX Plant microsatellite marker #339.

DE Plant microsatellite sequence; core repeat sequence; detection; probe;  
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;  
KW variety identification; genetic variability evaluation; primer; ss.

XX Eucalyptus grandis.

XX WO9967421-A1.

XX 29-DEC-1999.

XX 25-JUN-1999; 99WO-NZ00092.

XX 25-JUN-1998; 98US-0105307.

XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Havukkala IJ, Bloksberg LN, Glenn M;

DR WPI; 2000-116958/10.

XX New plant microsatellite markers and associated flanking species for  
PT the detection of polymorphic genetic markers -

XX Claim 1; Page 176-177; 392pp; English.

XX Sequences A31040-A32093 represent novel plant microsatellite sequences  
CC and associated flanking species. The sequences comprise a central core  
CC repeat sequence, especially selected from the sequences A32094-A32096  
CC with left and right flanking sequences. The polynucleotide sequences  
CC can be used in the detection of DNA polymorphisms, in genome mapping,  
CC in physical mapping, in positional cloning of genes, in variety  
CC identification and in evaluation of genetic variability within and  
CC between plant tissues, populations, cultivars, species and species  
CC groups. They may also be used to design hybridization probes for  
CC oligonucleotide fingerprinting and library screening and to design  
CC primers for microsatellite-primed PCR. Microsatellite markers are  
CC useful to locate specific economically useful genes in plant genomes.

XX Sequence 433 BP; 93 A; 108 C; 129 G; 102 T; 1 other;

Query Match  
Best Local Similarity 1.7%; Score 18; DB 21; Length 433;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 908 ttgtgtgtgtgtgagaga 925

Db 9 ttgtgtgtgtgtgagaga 26

RESULT 10  
N30074/C  
ID N30074 standard; DNA; 1325 BP.  
XX AC N30074;  
XX 25-APR-1992 (first entry)  
XX Sequence of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH)  
DE operon of S. cerevisiae transcription termination/polyadenylation  
DE region.

XX Yeast expression vector; promoter; Saccharomyces; ss.

XX Saccharomyces cerevisiae.

PH Key Location/Qualifiers

FT polyA\_signal 48..52

FT polyA\_signal /\*tag= a

FT polyA\_signal 315..319

FT polyA\_signal /\*tag= b

FT polyA\_signal 678..682

FT misc\_signal /\*tag= c

FT misc\_signal 21..23

FT misc\_signal /\*tag= d

FT misc\_signal 34..40

FT misc\_signal /\*tag= e

FT misc\_signal 163..165

FT misc\_signal /\*tag= f

FT misc\_signal 174..177

FT misc\_signal /\*tag= g

FT misc\_signal 179..181

FT misc\_signal /\*tag= h

FT misc\_signal 361..363

FT misc\_signal /\*tag= i

FT misc\_signal 389..392

FT misc\_signal /\*tag= j

FT misc\_signal 678..682

FT misc\_signal /\*tag= k

XX WO8304050-A.

XX 24-NOV-1983.

XX 19-MAY-1983; 33WO-2000712.

XX 19-MAY-1982; 32ND-0002091.

XX 19-MAY-1983; 33WO-EF00129.

XX (KONN ) GIST-BROCADES NV.

XX (HOLL/) HOLLEBERG C P.

XX (UNIL ) UNILEVER NV.

XX (OCTR-) INT OCTROOI MAATS.

XX Hollenberg C, De Leeuw A, Van Den Berg JA;

XX WPI; 1983-833205/48.

XX Kluyveromyces yeast cells - useful as hosts for recombinant DNA

XX material encoding for poly:peptide(s)

XX Example; Fig 6; 57pp; English.

XX The inventors claim a new cloning system which comprises a yeast of

XX the genus Kluyveromyces as a host. Suitable vectors are e.g.

XX vectors containing autonomously replicating sequences (ARS)

XX originating from Kluyveromyces of Saccharomyces. The example is for

XX the prepn. of Kluyveromyces SD11 lac4 trp1 expressing

XX preprothamatin and its various maturation forms after being

XX transformed with plasmid pURK 528-01 containing the structural gene

XX encoding proprothamatin, the KARS2 sequence from K. lactis, the

XX glyceraldehyde-3-phosphate dehydrogenase promoter from S. cerevisiae

CC and the TRP1 gene from *S. cerevisiae*.

XX Sequence 1325 BP; 376 A; 272 C; 232 G; 445 T; 0 other;

SQ Query Match 1.7%; Score 18; DB 4; Length 1325;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 753 ctcaaggagaaaaaaca 770

Db 167 CTCAGGAGAAAAAACA 150

RESULT 11

ID N40213/C

XX N40213 standard; DNA; 1325 BP.

AC N40213;

XX 29-NOV-1991 (first entry)

DE Transcription termination/polyadenylation region of glyceraldehyde  
DE 3-phosphate dehydrogenase operon.

XX Transcription terminator; polyadenylation; RNA polymerase II;  
KW Saccharomyces cerevisiae; plasmid; glyceraldehyde 3-phosphate hydrogenase.

XX Saccharomyces cerevisiae.

OS WO8404538-A.

PN 22-NOV-1984.

XX 21-MAY-1984; 84WO-EP00153.

XX 19-MAY-1983; 83NL-0200713.

XX (UNIL) UNILEVER NV.

XX Edens L, Russell SW, Visser C, Verrips CT;

PI WPI; 1984-300653/48.

XX DNA sequences to initiate transcription by yeast RNA polymerase -  
XX for expression to proteins in yeasts.  
PT Disclosure; Fig. 3; 81pp; English.

XX The DNA sequence is the transcription termination/polyadenylation  
XX region of the glyceraldehyde 3-phosphate dehydrogenase operon  
XX cloned into pRI 1-33. It forms part of a new DNA construct which  
XX can be inserted into a recombinant DNA plasmid or a yeast  
XX chromosome. This DNA construct comprises: a DNA sequence capable of  
XX initiating transcription by yeast RNA polymerase II (which includes  
XX at least part of the regulon gene of the glyceraldehyde 3-phosphate  
XX dehydrogenase genes of *Saccharomyces cerevisiae*); one or more structural  
XX genes different from the GAPDH genes of *S. cerevisiae* (asp. encoding  
XX thiamin or chymosin or their allelic or modified forms); and at least  
XX 2 of the following: one or more copies of the claimed sequence;  
XX one or more selection markers; either one or more nucleotide  
XX sequences allowing stable insertion into a yeast chromosome or one or  
XX more DNA sequences which regulated DNA replication in yeasts of genus  
XX *Saccharomyces* or *Kluyveromyces*; and a DNA sequence encoding a  
XX signal polypeptide of no more than 30 amino acids assisting the  
XX translocation of proteins. Yeasts contg. these constructs are  
XX cultivated for the prodn. of proteins.

SQ Sequence 1325 BP; 376 A; 270 C; 233 G; 446 T; 0 other;

Query Match 1.7%; Score 18; DB 5; Length 1325;  
Best Local Similarity 100.0%; Pred. No. 15;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 753 ctcaaggagaaaaaaca 770

Db 167 CTCAGGAGAAAAAACA 150

RESULT 12

XX X13303

XX X13303 standard; DNA; 1657 BP.

AC X13303;

XX 19-MAR-1999 (first entry)

XX Enterococcus faecalis genome contig SEQ ID NO:366.

XX Enterococcus faecalis; contig; detection; Enterococcal Infection;  
KW vaccine; attenuation; computer readable medium; ds.

XX Enterococcus faecalis.

OS WO9850555-A2.

PN 12-NOV-1998.

XX 04-MAY-1998; '98WO-US08985.

XX 14-NOV-1997; 97US-0066009.

XX 06-MAY-1997; 97US-0044031.

XX 16-MAY-1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Dillon PJ, Kunsch CA;

XX WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides  
XX - used to develop products for the detection of Enterococcus and for  
XX use in vaccines for prevention or attenuation of Enterococcus  
XX infection.

XX Claim 1; Page 1523-1524; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it  
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
XX X12938 to X13919 represent these nucleotide sequences which are primary  
XX nucleotide sequences, also known as contigs. The computer-based system  
XX can identify fragments of the Enterococcus faecalis genome with  
XX commercial importance. The products can be used to detect the presence  
XX of Enterococcus faecalis in samples. They can also be used for  
XX diagnosing Enterococcal infection in an animal and monitoring  
XX progression of disease, and for identifying agents which can be used to  
XX modulate the growth or pathogenicity of Enterococcus faecalis, or  
XX another related organism, in vivo or in vitro. In particular the  
XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
XX can be used in vaccines to prevent or attenuate an Enterococcal  
XX infection.

SQ Sequence 1657 BP; 547 A; 273 C; 342 G; 493 T; 2 other;

Query Match

Best Local Similarity 1.7%; Score 18; DB 20; Length 1657;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 165 gggaagaagagagagg 182

Db 1390 gggaagaagagagagg 1407

RESULT 13

N30016/c  
 ID N30016 standard; DNA; 2178 BP.  
 AC N30016;  
 XX  
 DT 25-APR-1992 (first entry)  
 DE Sequence of the promoter/regulation region of the glyceraldehyde-3  
 DE -phosphate dehydrogenase operon cloned in pF1 1-33.  
 DE  
 KW Yeast expression vector; promoter; Saccharomyces; ss.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PH Key Location/Qualifiers  
 FT TATA\_signal 78..83 /\*tag= a  
 FT TATA\_signal 241..244 /\*tag= b  
 FT TATA\_signal 717..720 /\*tag= c  
 FT TATA\_signal 721..724 /\*tag= d  
 FT TATA\_signal 900..904 /\*tag= e  
 FT TATA\_signal 1168..1172 /\*tag= f  
 FT TATA\_signal 1541..1545 /\*tag= g  
 FT terminator 217..219 /\*tag= h  
 FT terminator 224..227 /\*tag= i  
 FT terminator 233..235 /\*tag= j  
 FT terminator 525..528 /\*tag= k  
 FT terminator 546..548 /\*tag= l  
 FT terminator 643..645 /\*tag= m  
 FT terminator 647..649 /\*tag= n  
 FT terminator 656..659 /\*tag= o  
 FT terminator 662..664 /\*tag= p  
 FT terminator 668..670 /\*tag= q  
 FT terminator 695..697 /\*tag= r  
 FT terminator 874..876 /\*tag= s  
 FT terminator 887..893 /\*tag= t  
 FT terminator 1015..1017 /\*tag= u  
 FT terminator 1027..1030 /\*tag= v  
 FT terminator 1032..1034 /\*tag= w  
 FT terminator 1213..1215 /\*tag= x  
 FT terminator 1232..1234 /\*tag= y  
 FT terminator 1250..1252 /\*tag= z  
 FT misc\_feature 401..517 /\*tag= aa  
 XX  
 PN W08304051-A.  
 XX  
 PD 24-NOV-1983.

XX 19-MAY-1983; 33WO-EP00129.  
 XX  
 PR 19-MAY-1982; 32NL-0002091.  
 XX  
 PA (KONN ) GIST-BROCADES NV.  
 PA (HOLL/) HOLLEBERG C P.  
 PA (UNIL ) UNILEVER NV.  
 PA (OCTR-) INT OCTROOI MAATS.  
 XX  
 PI Edens L, Ledebøer A, Verrips CT, Van Den Berg JA;  
 XX  
 DR WPI; 1983-833207/48.  
 XX  
 PT Kluyveromyces yeast contg. recombinant DNA plasmid - for  
 FT cultivation in prodn. of preprothamatin or its modified forms  
 XX  
 PS Example; Fig 2; 57pp; English.  
 XX  
 CC The inventors claim a new cloning system which comprises a yeast of  
 CC the genus Kluyveromyces as a host. Suitable vectors are e.g.  
 CC vectors containing autonomously replicating sequences (ARS)  
 CC originating from Kluyveromyces of Saccharomyces. The example is for  
 CC the prepn. of Kluyveromyces SD11 lac4 trp1 expressing  
 CC preprothamatin and its various maturation forms after being  
 CC transformed with plasmid pURK 528-01 containing the structural gene  
 CC encoding preprothamatin, the KARS2 sequence from K. lactis, the  
 CC glyceraldehyde-3-phosphate dehydrogenase promoter from S. cerevisiae  
 CC and the TRP1 gene from S. cerevisiae.  
 XX  
 SQ Sequence 2178 BP; 639 A; 420 C; 389 G; 730 T; 0 other;  
 XX  
 Query Match 1.7%; Score 18; DB 4; Length 2178;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 753 ctcaagagaataaaaaa 770  
 DB 1020 CTCAGGAAATAAACA 1003  
 RESULT 14  
 V74953  
 ID V74953 standard; DNA; 2472 BP.  
 XX  
 AC V74953;  
 XX  
 DT 16-MAR-1999 (first entry)  
 XX  
 DE Staphylococcus aureus contig SEQ ID #642.  
 XX  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 XX  
 OS Staphylococcus aureus.  
 PH  
 FT Key Location/Qualifiers  
 FT misc\_feature 1441..1500 /\*tag= a  
 FT /\*note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence"  
 XX  
 PN EP786519-A2.  
 XX  
 PD 30-JUL-1997.  
 XX  
 PF 07-JAN-1997; 97EP-0100117.  
 XX



PR 05-JAN-1996; 9605-0009861.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 PI Rosen CA;  
 PI WPI; 1997-374922/35.  
 DR Polynucleotide(s) and proteins derived from *Staphylococcus aureus*  
 XX stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 XX  
 PS Claim 1; Page 1568-1569; 327ipp; English.  
 XX  
 XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 XX  
 XX Sequence 2472 BP; 840 A; 306 C; 399 G; 862 T; 65 other;

Query Match 1.7%; Score 18; DB 18; Length 2472;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 tatcattgttttcaatg 282  
 Db 2218 tatcattgttttcaatg 2235

RESULT 15  
 T74667/C  
 ID T74667 standard; DNA; 3505 BP.  
 XX  
 AC T74667;

XX 03-OCT-1997 (first entry)

XX Pig myogenin gene.

XX Myogenin; myogenesis; muscle growth; polymorphism; pig; ss.

XX Sus scrofa.

XX Key Location/Qualifiers

FT protein\_bind 22..71

FT /tag= a

FT /label= E1

FT /note= "E-box transcription factor binding

FT consensus sequence"

FT protein\_bind 210..215

FT /tag= b

FT /label= E2

FT /note= "E-box transcription factor binding

FT consensus sequence"

FT enhancer 456..464

FT /tag= c

FT /label= MEF-2  
 FT /note= "myocyte-specific enhancer factor 2"  
 FT 469...481  
 FT /tag= d  
 FT /label= NF-1  
 FT /note= "nuclear factor 1"  
 FT 494..498  
 FT /tag= e  
 FT 508...513  
 FT /tag= f  
 FT /label= E3  
 FT /note= "E-box transcription factor binding  
 FT consensus sequence"

FT 1..1042  
 FT /tag= g  
 FT /codon\_start= 572  
 FT /product= bHLH domain  
 FT 1043..1829  
 FT /tag= h  
 FT 1830..1911  
 FT /tag= i  
 FT /product= transactivation domain  
 FT 1912..2527  
 FT /tag= j  
 FT 2528..3505  
 FT /tag= k  
 FT /product= conserved C-terminal fragment  
 FT /note= "stop codon at 2647-2649"  
 FT 3594..3398  
 FT /tag= l

PN W09723644-A1.

XX 03-JUL-1997.

XX 20-DEC-1996; 96WO-NL00497.

XX 22-DEC-1995; 95EP-0203623.

XX (BEHE-) BEHEERMAATSCHAPPIJ VARKENSVERBETERING ZUID.

PA (COFO-) COFOR BV.

PA (DIER-) INST DIERHOUDERIJ EN DIERGEZONDHEID ID-DLO.

XX Soumillon AEM, Te Pas MFW;

XX WPI; 1997-351079/32.

DR P-FSDB; W22018.

XX Nucleic acid specific for the pig myogenin gene - for analysis of  
 PT association of PMG allele(s) with production traits, for use in  
 PT breeding programmes

XX Claim 1; Fig 1b; 42pp; English.

XX The pig myogenin gene (T74667) was isolated from porcine genomic DNA  
 CC by PCR amplification (see also T74668-71) and screening of a genomic  
 CC library with human myf cDNA fragments. It was assigned to  
 CC chromosome 9. MspI polymorphic sites were found in the promoter  
 CC region, in the second intron and at the 3' side of the gene. On  
 CC the basis of these polymorphisms, PCR-based methods are provided  
 CC (see also T74671-77) for identifying alleles of pig myogenin gene  
 CC that are associated with differences in production traits,  
 CC specifically those related to improved productivity, for use in  
 CC breeding programmes. Polymorphisms of the gene are related to  
 CC muscle development. Specific amplification of genomic fragments  
 CC allows many pigs to be genotyped rapidly.

XX Sequence 3505 BP; 777 A; 967 C; 1002 G; 758 T; 1 other;

Query Match 1.7%; Score 18; DB 18; Length 3505;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 908 ttgtgtgtgtgtgagaga 925  
|||||  
Db 2361 TTGTGTGTGTGTGAGAGA 2344

Search completed: November 4, 2000, 03:35:34  
Job time: 10998 sec





us-09-101-423a-2.olg.rni

Mon Nov 6 10:14:11 2000

```

;
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND
; ADDRESSEE: COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PATENT IN RELEASE #1.0, VERSION 1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/849,021
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/346,456
; FILING DATE: 28 NOVEMBER 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1064-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-992-7949
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-849-021-76

Query Match 1.6%; Score 17; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. NO. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 909 tgtgtgtgtgtgagaga 925
DB 18 TGTGTGTGTGTGAGAGA 2

RESULT 5
US-08-222-177A-312/C
; Sequence 312, Application US/08222177A
; Patent No. 5582979
; GENERAL INFORMATION:
; APPLICANT: Weber, James L.
; TITLE OF INVENTION: LENGTH POLYMORPHISMS IN
; TITLE OF INVENTION: (dG-da)n.(dG-dt)n SEQUENCES AND METHODS OF USING SAME
; NUMBER OF SEQUENCES: 460
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dewitt Ross & Stevens, S.C.
; STREET: 8000 Excelsior Drive, Suite 401
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,177A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,562
; FILING DATE: 21-APR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REGISTRATION NUMBER: 30,492
; REFERENCE/DOCKET NUMBER: 09865.601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 831-2100
; TELEFAX: (608) 831-2106
; TELEX:
; INFORMATION FOR SEQ ID NO: 379:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;

;
; APPLICATION NUMBER: US 07/341,562
; FILING DATE: 21-APR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REGISTRATION NUMBER: 30,492
; REFERENCE/DOCKET NUMBER: 09865.601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 831-2100
; TELEFAX: (608) 831-2106
; TELEX:
; INFORMATION FOR SEQ ID NO: 312:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: mfd97fs
; US-08-222-177A-312

Query Match 1.6%; Score 17; DB 1; Length 33;
Best Local Similarity 100.0%; Pred. NO. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 909 tgtgtgtgtgtgagaga 925
DB 22 TGTGTGTGTGTGAGAGA 6

RESULT 6
US-08-222-177A-379/C
; Sequence 379, Application US/08222177A
; Patent No. 5582979
; GENERAL INFORMATION:
; APPLICANT: Weter, James L.
; TITLE OF INVENTION: LENGTH POLYMORPHISMS IN
; TITLE OF INVENTION: (dG-da)n.(dG-dt)n SEQUENCES AND METHODS OF USING SAME
; NUMBER OF SEQUENCES: 460
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lewitt Ross & Stevens, S.C.
; STREET: 800C Excelsior Drive, Suite 401
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,177A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,562
; FILING DATE: 21-APR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REGISTRATION NUMBER: 30,492
; REFERENCE/DOCKET NUMBER: 09865.601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 831-2100
; TELEFAX: (608) 831-2106
; TELEX:
; INFORMATION FOR SEQ ID NO: 379:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;

```

```
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: mfd120rs
US-08-222-177A-379

Query Match          1.6%; Score 17; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 909 tgtgtgtgtgtgagaga 925
Db 21 TGTGTGTGTGTGAGAGA 5

RESULT 7
US-08-222-177A-65/c
; Sequence 65, Application US/08222177A
; Patent No. 5582979
; GENERAL INFORMATION:
; APPLICANT: Weber, James L.
; TITLE OF INVENTION: LENGTH POLYMORPHISMS IN
; NUMBER OF SEQUENCES: 460
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Demitt Ross & Stevens, S.C.
; STREET: 8000 Excelsior Drive, Suite 401
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,177A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,562
; FILING DATE: 21-APR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REGISTRATION NUMBER: 30,492
; REFERENCE/DOCKET NUMBER: 09865.601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 831-2100
; TELEFAX: (608) 831-2106
; TELEX:
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: mfd5rs
US-08-222-177A-65

Query Match          1.6%; Score 17; DB 1; Length 56;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 909 tgtgtgtgtgtgagaga 925
Db 22 TGTGTGTGTGTGAGAGA 6

RESULT 8
```

```
US-08-222-177A-244/c
; Sequence 244, Application US/08222177A
; Patent No. 5582979
; GENERAL INFORMATION:
; APPLICANT: Weber, James L.
; TITLE OF INVENTION: LENGTH POLYMORPHISMS IN
; NUMBER OF SEQUENCES: 460
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Demitt Ross & Stevens, S.C.
; STREET: 8000 Excelsior Drive, Suite 401
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53717-1314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,177A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,562
; FILING DATE: 21-APR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REGISTRATION NUMBER: 30,492
; REFERENCE/DOCKET NUMBER: 09865.601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 831-2100
; TELEFAX: (608) 831-2106
; TELEX:
; INFORMATION FOR SEQ ID NO: 244:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: mfd67rs
US-08-222-177A-244

Query Match          1.6%; Score 17; DB 1; Length 60;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 909 tgtgtgtgtgtgagaga 925
Db 35 TGTGTGTGTGTGAGAGA 19

RESULT 9
US-08-945-734-66
; Sequence 66, Application US/08945734
; Patent No. 6020130
; GENERAL INFORMATION:
; APPLICANT: LARRY GOLD
; APPLICANT: SUMEDHA JAYASENA
; TITLE OF INVENTION: NUCLEIC ACID LIGANDS
; TITLE OF INVENTION: THAT BIND TO AND INHIBIT DNA
; TITLE OF INVENTION: POLYMERASES
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson and Bratschun, L.L.C.
; STREET: 8400 East Prentice Avenue, Suite 200
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
```



;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/487,720  
;; FILING DATE: 7-JUNE-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/484,557  
;; FILING DATE: 7-JUNE-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Barry J. Swanson  
;; REGISTRATION NUMBER: 33,215  
;; REFERENCE/DOCKET NUMBER: NEX43C-US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (303) 793-3333  
;; TELEFAX: (303) 793-3433  
;; INFORMATION FOR SEQ ID NO: 70:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 78 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA  
US-08-945-734-70

Query Match 1.6%; Score 17; DB 5; Length 78;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 400 tggcggagcagaagcat 416  
Db 16 TGGCGGAGCAGAAGCAT 32

RESULT 12  
PCT-US96-09451-63  
;; Sequence 63, Application PC/TUS9609451  
;; GENERAL INFORMATION:  
;; APPLICANT: LARRY GOLD  
;; TITLE OF INVENTION: NUCLEIC ACID LIGAND INHIBITORS TO  
;; NUMBER OF SEQUENCES: 77  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Swanson and Bratschun, L.L.C.  
;; STREET: 8400 East Prentice Ave., Suite 200  
;; CITY: Denver  
;; STATE: Colorado  
;; COUNTRY: USA  
;; ZIP: 80111  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
;; MEDIUM TYPE: storage  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: MS-DOS  
;; SOFTWARE: WordPerfect 6.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US96/09451  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/487,426  
;; FILING DATE: 7-JUNE-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/487,720  
;; FILING DATE: 7-JUNE-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/484,557  
;; FILING DATE: 7-JUNE-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Barry J. Swanson  
;; REGISTRATION NUMBER: 33,215  
;; REFERENCE/DOCKET NUMBER: NEX43C/PCT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (303) 793-3333

;; TELEFAX: (303) 793-3433  
;; INFORMATION FOR SEQ ID NO: 63:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 78 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA  
PCT-US96-09451-63

Query Match 1.6%; Score 17; DB 6; Length 78;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 400 tggcggagcagaagcat 416  
Db 16 TGGCGGAGCAGAAGCAT 32

RESULT 13  
PCT-US96-09451-65  
;; Sequence 65, Application PC/TUS9609451  
;; GENERAL INFORMATION:  
;; APPLICANT: LARRY GOLD  
;; TITLE OF INVENTION: NUCLEIC ACID LIGAND INHIBITORS TO  
;; NUMBER OF SEQUENCES: 77  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Swanson and Bratschun, L.L.C.  
;; STREET: 8400 East Prentice Ave., Suite 200  
;; CITY: Denver  
;; STATE: Colorado  
;; COUNTRY: USA  
;; ZIP: 80111  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb  
;; MEDIUM TYPE: storage  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: MS-DOS  
;; SOFTWARE: WordPerfect 6.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US96/09451  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/487,426  
;; FILING DATE: 7-JUNE-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/487,720  
;; FILING DATE: 7-JUNE-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/484,557  
;; FILING DATE: 7-JUNE-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Barry J. Swanson  
;; REGISTRATION NUMBER: 33,215  
;; REFERENCE/DOCKET NUMBER: NEX43C/PCT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (303) 793-3333  
;; TELEFAX: (303) 793-3433  
;; INFORMATION FOR SEQ ID NO: 65:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 78 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA  
PCT-US96-09451-65

Query Match 1.6%; Score 17; DB 6; Length 78;



```

GENERAL INFORMATION:
APPLICANT: Weber, James L.
TITLE OF INVENTION: LENGTH POLYMORPHISMS IN
TITLE OF INVENTION: (GC-GA)n.(AG-GT)n SEQUENCES AND METHODS OF USING SAME
NUMBER OF SEQUENCES: 460
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lewitt Ross & Stevens, S.C.
STREET: 800C Excelsior Drive, Suite 401
CITY: Madison
STATE: Wisconsin
COUNTRY: USA
ZIP: 53717-1914
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IEM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,177A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/341,562
FILING DATE: 21-APR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Sara, Charles S.
REGISTRATION NUMBER: 30,492
REFERENCE/DOCKET NUMBER: 09865.601
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 831-2100
TELEFAX: (608) 831-2106
TELEX:
INFORMATION FOR SEQ ID NO: 166:
SEQUENCE CHARACTERISTICS:
LENGTH: 91 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: mfd39rs
US-08-222-177A-166
Query Match 1.6% Score 17 DB 1 Length 91;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 909 tgtgtgtgtgtgagaga 925
Db 63 TGTGTGTGTGTGAGAGA 47
Search completed: November 4, 2000, 03:31:32
Job time: 11882 sec

```

search completed: Ncvember 4, 2000, 03:31:32  
Job time: 11882 sec:

```
Best Local Similarity 100.0%; Pried. No. 26;
Matches 17; Conservative 0; Mismatches 0; Gaps 0;

QY 400 tggcggagcagaacat 416
      |||||
Db 16 TGGCGGAGCAGAACAT 32

RESULT 14
PCT-US96-09451-66
; Sequence 66, Application PC/TUS9609451
; GENERAL INFORMATION:
; APPLICANT: LARRY RAY
; APPLICANT: SUMEDHA JAYASENA
; TITLE OF INVENTION: NUCLEIC ACID LIGAND INHIBITORS TO
; TITLE OF INVENTION: DNA POLYMERASES
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson and Bratschun, L.L.C.
; STREET: 8400 East Prentice Ave., Suite 200
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,426
; FILING DATE: 7-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,720
; FILING DATE: 7-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/484,557
; FILING DATE: 7-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX43C/PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
PCT-US96-09451-66
```

```

Query Match      1.6%; Score 17; DB 6; Length 78;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 15  
US-08-222-177A-166/c  
; Sequence 166, Application US/08222177A  
; Patent No. 5582979

Mon Nov 6 10:14:11 2000

us-09-101-423a-2.olig.rni

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 01:08:36 ; Search time 3299.68 Seconds  
(without alignments)  
1982.437 Million cell updates/sec

Title: US-09-101-423A-2  
Perfect score: 1058  
Sequence: 1 attgtgtgagcctattagc.....acacattogaacgttggtc 1058

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 7189864 seqs, 3091403243 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estom:\*  
68: em\_estov1:\*  
69: em\_estov2:\*  
70: em\_estpl1:\*  
71: em\_estpl2:\*  
72: em\_estpl3:\*  
73: em\_estpl4:\*  
74: em\_estpl5:\*  
75: em\_estrol:\*  
76: em\_estro2:\*  
77: em\_estro3:\*  
78: em\_estro4:\*  
79: em\_estro5:\*  
80: em\_estro6:\*  
81: em\_estro7:\*  
82: em\_estro8:\*  
83: em\_estro9:\*  
84: em\_estro10:\*  
85: em\_estro11:\*  
86: em\_estro12:\*  
87: em\_estro13:\*  
88: gb\_gss1:\*  
89: gb\_gss2:\*  
90: gb\_gss3:\*  
91: gb\_gss4:\*  
92: em\_gss1:\*  
93: em\_gss2:\*  
94: em\_gss3:\*  
95: em\_gss4:\*  
96: gb\_gss5:\*  
97: gb\_gss6:\*  
98: gb\_gss7:\*  
99: gb\_gss8:\*  
100: gb\_gss9:\*  
101: em\_gss5:\*  
102: em\_gss6:\*  
103: em\_gss7:\*  
104: em\_gss8:\*  
105: em\_gss9:\*  
106: em\_gss10:\*  
107: em\_gss11:\*  
108: gb\_gss10:\*  
109: gb\_gss11:\*  
110: em\_gss12:\*  
111: gb\_gss12:\*  
112: gb\_gss13:\*  
113: gb\_gss14:\*  
114: gb\_gss15:\*  
115: gb\_gss16:\*  
116: gb\_gss17:\*

117: gb\_gss18:\*  
 118: gb\_gss19:\*  
 119: em\_gss13:\*  
 120: gb\_gss20:\*  
 121: gb\_gss21:\*  
 122: gb\_gss22:\*  
 123: gb\_gss23:\*  
 124: gb\_gss24:\*  
 125: em\_gss14:\*  
 126: em\_gss15:\*  
 127: em\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
C 1	42	4.0	384	114	AZ040211	RPCI-23-3
C 2	25	2.4	311	10	AI425066	tg50h03.x
C 3	21	2.0	843	123	CNS03012	Tetraodon
C 4	21	2.0	1041	122	CNS02HPB	AI197912 Tetraodon
C 5	20	1.9	112	1	AA107788	mo48b09.r
C 6	20	1.9	195	26	BB063640	BB063640
C 7	20	1.9	202	24	AW841795	AW841795 RC3-CN001
C 8	20	1.9	468	108	AQ716924	HS_5424_B
C 9	20	1.9	515	114	AQ058019	RPCI-23-4
C 10	20	1.9	528	113	AQ984351	RPCI-23-3
C 11	20	1.9	552	97	AQ406252	HS_3059_A
C 12	20	1.9	564	99	AQ596093	HS_2132_B
C 13	20	1.9	619	124	FR0002515	Z86298 F.rubripes
C 14	20	1.9	711	96	AQ392089	CITR1-EI-
C 15	20	1.9	957	122	CNS0279B	AI184376 Tetraodon
C 16	20	1.9	1012	123	CNS040YO	AL269529 Tetraodon
C 17	19	1.8	160	24	AW741643	AW741643 ur58b08.x
C 18	19	1.8	167	40	W84695	W84695 zh52a02.sl
C 19	19	1.8	182	32	BE009502	PM1-BN017
C 20	19	1.8	188	113	AQ007244	RPCI-23-3
C 21	19	1.8	223	11	AI618327	AI618327 zehs0093.s
C 22	19	1.8	231	27	BB452103	BB452103 BB152103
C 23	19	1.8	232	7	AA998972	UI-R-CO-1
C 24	19	1.8	265	12	AI631492	AI631492 wa89f08.x
C 25	19	1.8	309	39	T23451	seg2965 1-N
C 26	19	1.8	313	14	AJ728098	4A3A-ABA-
C 27	19	1.8	315	6	AA740192	ob26f03.s
C 28	19	1.8	322	31	BB451276	BB451276
C 29	19	1.8	323	31	BB458836	BB458836
C 30	19	1.8	327	117	AZ230299	RPCI-23-6
C 31	19	1.8	336	38	N67359	Yz50e06.sl
C 32	19	1.8	356	12	AI660888	wf21a05.x
C 33	19	1.8	386	120	B89007	CIT-HSP-217
C 34	19	1.8	391	37	H00500	Y724Q04.r1
C 35	19	1.8	401	38	RA1744	Yg1ld06.sl
C 36	19	1.8	402	112	AQ928271	RPCI-23-2
C 37	19	1.8	408	37	H46142	Yol3C11.sl
C 38	19	1.8	409	3	AA333545	AA333545
C 39	19	1.8	410	12	AI640657	AI640657 tz69906.x
C 40	19	1.8	414	14	AJ7281786	4A3A-AAK-
C 41	19	1.8	415	6	AA769156	nz12e05.s
C 42	19	1.8	417	100	AQ616952	HS_5156_A
C 43	19	1.8	423	4	AA416573	znl8c09.s
C 44	19	1.8	432	4	AA489740	aa43c06.r
C 45	19	1.8	435	122	CNS02JMU	AL200415 Tetraodon

## ALIGNMENTS

RESULT 1  
 AZ040211/c

LOCUS AZ040211 384 bp DNA GSS 01-MAR-2000  
 DEFINITION RPCI-23-367G16.TV RPCI-23 Mus musculus genomic clone RPCI-23-367G16  
 , DNA sequence.  
 ACCESSION AZ040211  
 VERSION AZ040211.1 GI:7131635  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE 1 (bases 1 to 384)  
 AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret  
 B., Levin, M., McGann, S., Isegaye, G., Geer, K., Kroll, M., de Jong, P.  
 and Fraser, C.M.  
 TITLE Mouse BAC End Sequences from Library RPCI-23  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: RPCI-23-367G16.TJ  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301.838 0200  
 Fax: 301.838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
 or from Resea ch Genetics (info@resgen.com). BAC end page:  
 http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 357 row: G column: 16  
 Seq primer: 77  
 Class: BAC ends.  
 Location/Qualifiers  
 1. 384  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-367G16"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site: 1:  
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBACE3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 105 a 86 c 86 g 107 t  
 ORIGIN  
 Query Match 4.0%; Score 42; DB 114; Length 384;  
 Best Local Similarity 100.0%; Pred.No. 1.9e-11;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 426 ttctgataaa-agtactttgtaatatatttgagacggagct 467  
 |||||  
 Db 369 TTCTGATAAAAGTACTTTGTAATATTTGGAGACGGGAGCT 328  
 RESULT 2  
 LOCUS AI425066 311 bp mRNA EST 30-MAR-1999  
 DEFINITION tg50h03.1 Soares\_NFL\_T\_GBC.S1 Homo sapiens cDNA clone  
 IMAGE:2112245 3', mRNA sequence.  
 ACCESSION AI425066  
 VERSION AI425066.1 GI:4270997  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Tetraodon nigroviridis DNA sequence

Unpublished  
3 (bases 1 to 843)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/tetraodon.  
Location/Qualifiers  
1. 843  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="041M17"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C05G041AG09SP1-end :  
PUC-ori"  
BASE COUNT 207 a 177 c 196 g 251 t 12 others  
ORIGIN

Query Match 2.0%; Score 21; DB 123; Length 843;  
Best Local Similarity 100.0%; Pred. No. 4.4; 0; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 0;

QY 535 acaggaccacccaaacacgctgac 555  
|||||  
Db 415 ACAGGACCCCAACACGCTGAC 435  
|||||

RESULT 4  
CNS02HPB/c 1041 bp DNA GSS 13-MAY-2000  
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone  
DEFINITION 140C05 of library G from Tetraodon nigroviridis, genomic survey  
sequence;  
AL197912  
VERSION AL197912.1 GI:7836062  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;  
Holacanthopterygii; Acanthopterygii; Percomorpha;  
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.  
1 (bases 1 to 1041)  
Roest-Crollius.H., Jaillon.O., Dasilva.C., Fizames.C., Fisher.C.,  
Bouneau.L., Billault.A., Quetier.F., Saurin.W., Bernot.A. and  
Weissenbach.J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 1041)  
Roest-Crollius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C.,  
Bernot.A., Fizames.C., Wincker.P., Brottier.P., Quetier.F.,  
Saurin.W. and Weissenbach.J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 1041)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
Location/Qualifiers  
1. 1041  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
1 (bases 1 to 311)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 382 Std Error: 0.00  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. 311  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2112245"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with  
a modified polylinker; Site: 1; Not 1; Site: 2; Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbHL19W, testis NHT, and B-cell  
NCL-CGAP GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo.  
BASE COUNT 97 a 62 c 55 g 97 t  
ORIGIN

Query Match 2.4%; Score 25; DB 10; Length 311;  
Best Local Similarity 100.0%; Pred. No. 0.028;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 ttcatgtctaccatattgcctgac 398  
|||||  
Db 169 TTCATGCTGCTACCATATGCTGAC 193  
|||||

RESULT 3  
CNS0301Z 843 bp DNA GSS 17-MAY-2000  
LOCUS Tetraodon nigroviridis genome survey sequence PUC-ori end of clone  
DEFINITION 041M17 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
AL252800  
VERSION AL252800.1 GI:7973812  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;  
Holacanthopterygii; Acanthopterygii; Percomorpha;  
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.  
1 (bases 1 to 843)  
Roest-Crollius.H., Jaillon.O., Dasilva.C., Fizames.C., Fisher.C.,  
Bouneau.L., Billault.A., Quetier.F., Saurin.W., Bernot.A. and  
Weissenbach.J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 843)  
Roest-Crollius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C.,  
Bernot.A., Fizames.C., Wincker.P., Brottier.P., Quetier.F.,  
Saurin.W. and Weissenbach.J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 843)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
Location/Qualifiers  
1. 843  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"



transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATTAATCCGCCGCCGCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 46 a 52 c 40 g 56 t 1 others

ORIGIN

Query Match 1.9%: Score 20; DB 26; Length 195;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 909 tctgtgtgtgtgagagacaa 928  
|||||

Db 147 TGTGTGTGTGTGAGAGACAA 128

RESULT 7  
LOCUS AW841795 202 bp mRNA EST 18-MAY-2000  
DEFINITION RC3-CN0018-120200-014-g09\_1 CN0018 Homo sapiens cDNA, mRNA  
ACCESSION AW841795  
VERSION AW841795.1 GI:7935778  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 202)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC3-CN0018-120-014-g09.1&t3=2000-02-12&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 202.  
Location/Qualifiers  
1. .202  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CN0018"  
/dev\_stage="Adult"  
/note="Organ: colon\_normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 64 a 38 c 44 g 56 t

ORIGIN

Query Match 1.9%: Score 20; DB 24; Length 202;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 906 acttgtgtgtgtgtgagaga 925  
|||||

Db 143 ACTTGTGTGTGTGTGAGAGA 162

RESULT 8  
LOCUS AQ716924/c 468 bp DNA GSS 13-JUL-1999  
DEFINITION HS\_5424\_B2\_E10\_SP6E\_RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-1000 Col-20 Row-J, DNA sequence.  
ACCESSION AQ716924  
VERSION AQ716924.1 GI:5466240  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 468)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@delong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Research Genetics (Info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
Plate: 1000 row: J column: 20  
Seq primer: SP6  
Class: EAC ends  
High quality sequence stop: 468.  
Location/Qualifiers  
1. .468  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Plate-1000 Col-20 Row-J"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 123 a 119 c 102 g 123 t 1 others

ORIGIN

Query Match 1.9%: Score 20; DB 108; Length 468;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 784 attatttgagatgaagtct 803  
|||||

Db 345 ATTATTGGAGATGAAGTCT 326

RESULT 9  
LOCUS AQ716924/c 468 bp DNA GSS 13-JUL-1999  
DEFINITION HS\_5424\_B2\_E10\_SP6E\_RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-1000 Col-20 Row-J, DNA sequence.  
ACCESSION AQ716924  
VERSION AQ716924.1 GI:5466240  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 468)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@delong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Research Genetics (Info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
Plate: 1000 row: J column: 20  
Seq primer: SP6  
Class: EAC ends  
High quality sequence stop: 468.  
Location/Qualifiers  
1. .468  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Plate-1000 Col-20 Row-J"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 123 a 119 c 102 g 123 t 1 others

ORIGIN

Query Match 1.9%: Score 20; DB 108; Length 468;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 784 attatttgagatgaagtct 803  
|||||

Db 345 ATTATTGGAGATGAAGTCT 326

RESULT 9  
LOCUS AQ716924/c 468 bp DNA GSS 13-JUL-1999  
DEFINITION HS\_5424\_B2\_E10\_SP6E\_RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-1000 Col-20 Row-J, DNA sequence.  
ACCESSION AQ716924  
VERSION AQ716924.1 GI:5466240  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 468)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@delong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Research Genetics (Info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
Plate: 1000 row: J column: 20  
Seq primer: SP6  
Class: EAC ends  
High quality sequence stop: 468.  
Location/Qualifiers  
1. .468  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Plate-1000 Col-20 Row-J"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 123 a 119 c 102 g 123 t 1 others

ORIGIN

AZ058019  
LOCUS  
DEFINITION RPCI-23-427J21.TV RPCI-23 Mus musculus genomic clone RPCI-23-427J21  
, DNA sequence.  
ACCESSION AZ058019 515 bp DNA GSS 30-MAR-2000  
VERSION AZ058019.1 GI:7349256  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 515)  
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other\_GSSs: RPCI-23-427J21.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics ([http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html))  
Plate: 427 row: J' column: 21  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1..515  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-427J21"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1: EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 120 a 95 c 127 g 173 t  
ORIGIN  
Query Match 1.9%; Score 20; DB 114; Length 515;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 901 tgtacacttggtgtgtgtg 920  
|||||  
Db 333 TGTACTTGTGTGTGTG 352

RESULT 10  
AQ984351/c  
LOCUS  
DEFINITION RPCI-23-323M16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-323M16  
, DNA sequence.  
ACCESSION AQ984351  
VERSION AQ984351.1 GI:6817556  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 528)  
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics ([http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html))  
Plate: 323 row: M column: 16  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1..528  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-323M16"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1: EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 135 a 75 c 71 g 245 t  
ORIGIN  
Query Match 1.9%; Score 20; DB 113; Length 528;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 757 aggaataaaacacacatta 776  
|||||  
Db 30 AGGAAAAAACAACACATTA 11

RESULT 11  
AQ406252  
LOCUS  
DEFINITION HS\_5059\_A2\_B11\_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-635 Col-22 Row-C, DNA sequence.  
ACCESSION AQ406252  
VERSION AQ406252.1 GI:4411617  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

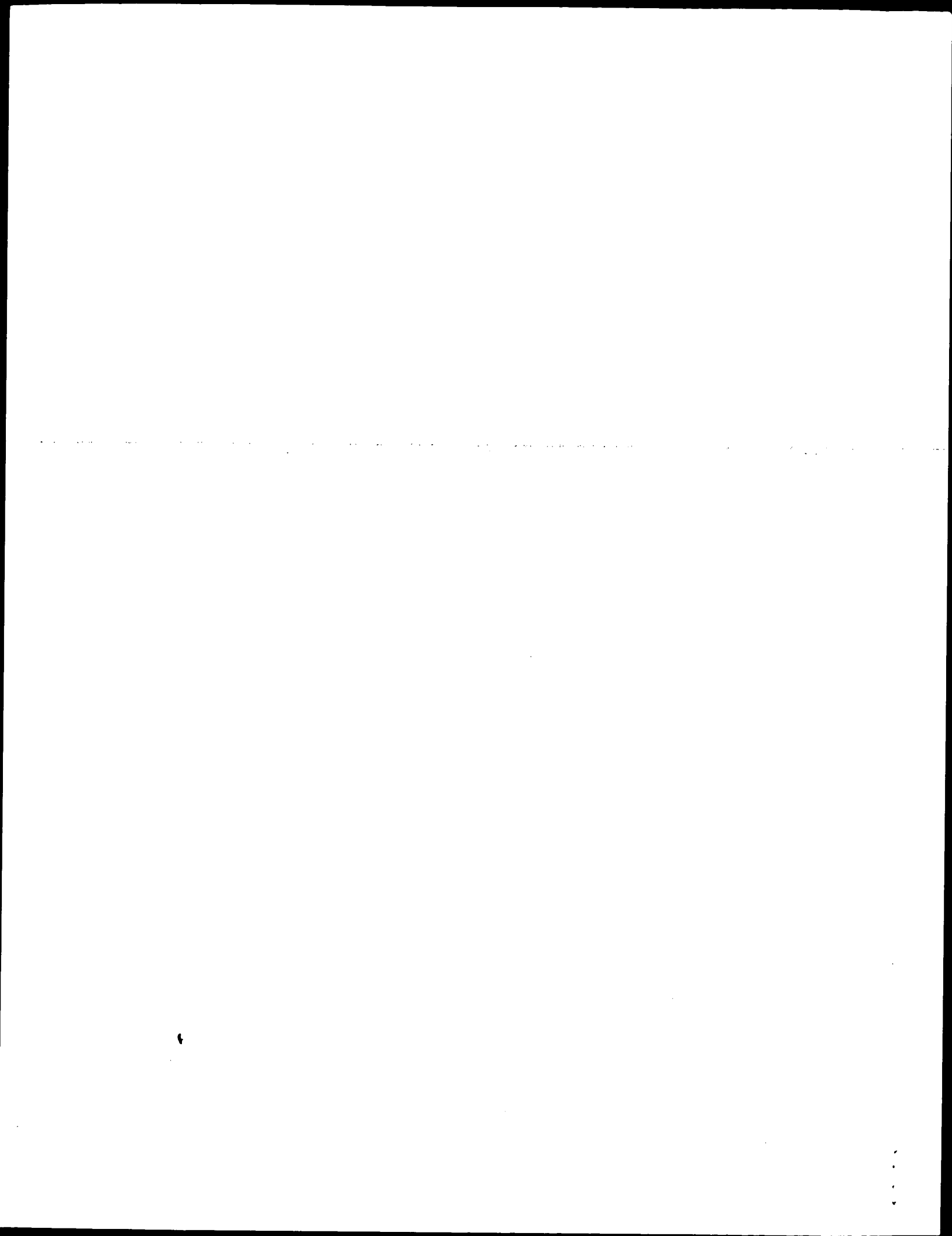
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 552)  
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.  
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
93380589











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 4, 2000, 00:22:45 ; Search time 7109.34 Seconds  
(without alignments)  
649.976 Million cell updates/sec

Title: US-09-101-423A-2  
Perfect score: 1058  
Sequence: 1 attgtgtgagcctattagc.....acacattggaacgttggtc 1058

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues  
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : GenEmbl.\*
- 1: gb\_bal.\*
  - 2: gb\_bal.\*
  - 3: gb\_bal.\*
  - 4: gb\_bal.\*
  - 5: gb\_bal.\*
  - 6: gb\_bal.\*
  - 7: gb\_bal.\*
  - 8: gb\_bal.\*
  - 9: gb\_bal.\*
  - 10: gb\_bal.\*
  - 11: gb\_bal.\*
  - 12: gb\_bal.\*
  - 13: gb\_bal.\*
  - 14: gb\_bal.\*
  - 15: gb\_bal.\*
  - 16: gb\_bal.\*
  - 17: gb\_bal.\*
  - 18: gb\_bal.\*
  - 19: gb\_bal.\*
  - 20: gb\_bal.\*
  - 21: gb\_bal.\*
  - 22: gb\_bal.\*
  - 23: gb\_bal.\*
  - 24: gb\_bal.\*
  - 25: gb\_bal.\*
  - 26: gb\_bal.\*
  - 27: gb\_bal.\*
  - 28: gb\_bal.\*
  - 29: gb\_bal.\*
  - 30: gb\_bal.\*
  - 31: gb\_bal.\*
  - 32: gb\_bal.\*
  - 33: gb\_bal.\*
  - 34: gb\_bal.\*
  - 35: gb\_bal.\*
  - 36: gb\_bal.\*
  - 37: gb\_bal.\*
  - 38: gb\_bal.\*
  - 39: gb\_bal.\*
  - 40: gb\_bal.\*
  - 41: gb\_bal.\*
  - 42: gb\_bal.\*
  - 43: gb\_bal.\*

- 44: em\_hum5.\*
- 45: gb\_pl3.\*
- 46: gb\_pr5.\*
- 47: gb\_htg8.\*
- 48: gb\_htg9.\*
- 49: gb\_htg10.\*
- 50: gb\_htg11.\*
- 51: gb\_htg12.\*
- 52: gb\_htg13.\*
- 53: gb\_htg14.\*
- 54: gb\_in3.\*
- 55: gb\_htg15.\*
- 56: gb\_htg16.\*
- 57: gb\_htg17.\*
- 58: em\_htg4.\*
- 59: em\_htg5.\*
- 60: em\_htg6.\*
- 61: em\_htg7.\*
- 62: em\_hum6.\*
- 63: gb\_htg18.\*
- 64: gb\_htg19.\*
- 65: gb\_bal.\*
- 66: em\_htg8.\*
- 67: em\_htg9.\*
- 68: em\_htg10.\*
- 69: em\_htg11.\*
- 70: em\_htg12.\*
- 71: em\_htg13.\*
- 72: em\_htg14.\*
- 73: em\_htg15.\*
- 74: em\_htg16.\*
- 75: em\_htg17.\*
- 76: em\_htg18.\*
- 77: em\_htg19.\*
- 78: em\_htg20.\*
- 79: em\_htg21.\*
- 80: em\_htg22.\*
- 81: em\_htg23.\*
- 82: gb\_pr6.\*
- 83: gb\_pr7.\*
- 84: gb\_htg20.\*
- 85: gb\_htg21.\*
- 86: gb\_htg22.\*
- 87: gb\_htg23.\*
- 88: gb\_ro.\*
- 89: gb\_sts1.\*
- 90: gb\_sts2.\*
- 91: gb\_sy.\*
- 92: gb\_un.\*
- 93: gb\_vil.\*
- 94: gb\_vil.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1058	100.0	1058	5	A64059	A64059 Sequence 2
C 2	317.6	30.0	152402	48	AC022658	AC022658 Homo sapi
C 3	317.6	30.0	163317	35	CNS010X0	AL138995 Human chr
C 4	317.6	30.0	195713	87	CNS050BP	AL163952 Homo sapi
C 5	56.4	5.3	203111	86	AP001647	AP001647 Homo sapi
C 6	56.4	5.3	203112	53	AC040896	AC040896 Homo sapi
C 7	54.8	5.2	182540	37	AC012056	AC012056 Homo sapi
C 8	54.6	5.2	87795	49	AC023458	AC023458 Homo sapi
C 9	54.6	5.2	110000	36	AC010909_0	AC010909 Homo sapi
C 10	54.6	5.2	206177	10	AC008072	AC008072 Homo sapi
C 11	53.8	5.1	167095	35	CNS01RHA	AL161666 Human chr
C 12	53.8	5.1	181905	87	CNS01RIB	AL163052 Homo sapi

c 13 53.8 5.1 209706 55 AC067885  
 14 52.4 5.0 184511 56 AC069437  
 15 52.2 4.9 56804 46 HS77N19  
 16 52.2 4.9 176197 39 AC016712  
 17 52.2 4.9 192913 48 AC023100  
 18 52 4.9 133951 63 AL139045  
 19 51.8 4.9 11444 82 HSL96F8  
 20 51.8 4.9 11444 82 HSL96F8A  
 21 51.8 4.9 29408 28 AC009804  
 22 51.8 4.9 178446 85 AC009804  
 23 51.6 4.9 98209 46 HS260J9  
 24 51.6 4.9 134394 55 AC067916  
 25 51.4 4.9 337101 83 HSXSRPXR  
 26 51.2 4.8 136598 35 CWS01DT6  
 27 50.8 4.8 59461 37 AC011889  
 28 50.8 4.8 107626 11 AL133334  
 29 50.8 4.8 163364 40 AC018588  
 30 50.8 4.8 167585 48 AC023013  
 31 50.8 4.8 174324 85 AL360272  
 32 50.8 4.8 184743 82 HSAF003627  
 33 50.8 4.8 203702 85 AL357498  
 34 50.8 4.8 205302 40 AC018885  
 35 50.8 4.8 207620 35 AP001331  
 36 50.8 4.8 244499 64 AL161905  
 37 50.6 4.8 107574 28 AC008950  
 38 50.6 4.8 137557 10 AC009414  
 39 50.6 4.8 159044 50 AC025269  
 40 50.6 4.8 171243 57 AC074350  
 41 50.6 4.8 184675 39 AC016602  
 42 50.4 4.8 149258 53 AC044787  
 43 50.4 4.8 196596 47 AC020688  
 44 50.2 4.7 148192 47 AC021507  
 45 50.2 4.7 158309 56 AC073158

## ALIGNMENTS

RESULT 1  
 A64059 1058 bp DNA PAT 29-MAR-1999  
 LOCUS Sequence 2 from Patent WO9725443.  
 DEFINITION A64059  
 ACCESSION A64059.1 GI:3717533  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCES  
 1 (bases 1 to 1058)  
 Rudland, P. S. and Barraclough, B. R.  
 TITLE METASTASIS INDUCING DNA'S  
 JOURNAL Patent: WO 9725443-A 17-JUL-1997;  
 UNIV LIVERPOOL (GB)  
 FEATURES  
 Location/Qualifiers  
 1..1058  
 /organism="unidentified"  
 /db\_xref="taxon:32844"  
 BASE COUNT 320 a 219 c 249 g 270 t  
 ORIGIN

Query Match 100.0%; Score 1058; DB 5; Length 1058;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-279;  
 Matches 1058; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgctgtgagcctattagcgacatttggtagcgccttttaagggggtagatacaag 60  
 Db 1 ATTGCTGTGACCTATTAGCGACATTGTGAGCGCCCTTTAAGGGGTAGATACAAAG 60  
 Qy 61 aatgggtgaaattctgtgcacaaacgtctccatgttttcaacaattacactgcaacc 120  
 Db 61 AATGGGTGAAATTCTGTGCGCAAAACGGCTCTCCATGTTTTCACAAATTACACTTGAACC 120

Qy 121 tgtgtgcagcagcagcaatttagggatgtatgggacaggttcgggaaagagaggaag 180  
 Db 121 TGTGTCAGCAGCGCAGCAATTTAGGGATGTATGGGACAGGTCGGGGAAGAGGAGAAG 180  
 Qy 181 ggtaaagaaagcagcagcagcgttaaaagccaaacagctccaggagatctctgtagaata 240  
 Db 181 GGTAAAGAAAGACAGCAGCAGCTTAAAGTCCAAACAGCTCCAGGAGATATCTGTAGAATA 240  
 Qy 241 acatcagacattgaggaattgatcatctgtttttcaatgggtatcccaagggaact 300  
 Db 241 ACATCAGACCATGAGGAAATTCATATCATTTTCAATGGGTATGCCAAGGGAAT 300  
 Qy 301 ttcacatgttataaataattactgtgacataatccaattggaaattgccccacaca 360  
 Db 301 TTCATCTGATTAATAATAATTAATCTGCTGCACTAAATCCAATTGGAAATGCCCAACACA 360  
 Qy 361 attattctccactctactgctaccatgctgacgtacgtgacgtgacgtgacgtacgt 420  
 Db 361 ATTATTCTTCACCTTCATGCTGCTACCATATGCTGACGTGACGTGCGGAGCAAGCATTTCC 420  
 Qy 421 tccggtctgataaataagtagtactttgtataattttggagacgggagctctgtgacagga 480  
 Db 421 TCCGCTCTGATAAATAAGTAGTACTTTGTAAATATTGGAGACGGGAGCTCTGGTGACAGGA 480  
 Qy 481 acagctacaaacgggctctttatcatgctcccgatagaggccctctttgacgtacagga 540  
 Db 481 ACAGCTACAAACGGGCTCTTTATCATGTTCCCGATAGAGGCCCTCTTTAGCGTACAGGA 540  
 Qy 541 ccccaaaacajtcaggatgctgtgaatttccctccatgaagcctgtgttcacaaattagcaa 600  
 Db 541 CCCCCAAACAGTCAGGATGCTGTGAATTTCCATGAAGCCTTTGTCACAAATTAGCAA 600  
 Qy 601 ccattggaggaagcaggtgcactgtctaccacaaagtgagcactttccaaagagcacacat 660  
 Db 601 CCATTGGAGGAAGCAGGCTGCACCTGTCTACCAACAAGTGGCACTTTTCCAAAGAGCACACAT 660  
 Qy 661 atattggagcagacattttgctgctgctgctgctgctgctgctgctgctgctgctgct 720  
 Db 661 ATATTGGAGCAAGACATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 Qy 721 ttattaaactggctttttttttgaacacccactcaagggaaaaaaacacacacttaggt 780  
 Db 721 TTTATTAACCTGGCTTTTCTTTTGAACACCCCACTCAAGGAAAAAACAACACACTTAGGT 780  
 Qy 781 gacattatttgagatgaagcttttatagagatgcttaagtttaaacagagacttttaag 840  
 Db 781 GACATTATTGGAGATGAAGTCTTTTATAGATGCTTAAGTTTAAACGAGACTTTTAAAG 840  
 Qy 841 ccggtctctatccatttaataatggtgtccctacaaaggaagcgtgagggagcagctacaag 900  
 Db 841 CCGGCTCTATCCATTTAATGAATGGTGTCCCTACAAAGGAAGAAACTGGGACAGAGGTA 900  
 Qy 901 tctacactgt 960  
 Db 901 TGTACACTTGT 960  
 Qy 961 agagaagggcgacccctatttcacacgtgacacacacacacacacacacacacacacacac 1020  
 Db 961 AGAGAAAGGCGACCCCTATTTCACACTGAGCAAAACCACTGATGTGTGGTGTGATAGTA 1020  
 Qy 1021 gattatcccccaagactcacacattccgaacgttggtc 1058  
 Db 1021 GAGTATCCCCCAAGACTCACACATTCGAACGCTTGGTC 1058

RESULT 2  
 AC022658/c  
 LOCUS  
 DEFINITION Homo sapiens clone RP11-29G17, WORKING DRAFT SEQUENCE, 9 unordered  
 pieces.  
 AC022658  
 AC022658.3 GI:7249113  
 VERSION  
 KEYWORDS HTG; HTG\_PHASE1; HTGS\_DRAFT.



QY	563	tgaatttcottccatgaagccttgtttacaaattagcaaccattggaagcaagcagctgca	622
Db	40201	AGAATTCCTTTCCCAACATTTCTTGTTCAGAAATTAGACGAATTTGGAGGAGCATGTTGGG	40142
QY	623	ctgtctaccacaagtggcactttccaaagagcacacatatattggagcaagacatttgc	682
Db	40141	CACTCT--ACATGAGCCATGCTGCCCCNAGGGCGCATACATCAGTGAGACATTTTGC	40085
QY	683	tggctgacttggctgtgtgtaagctgat--aaactgctctatttattaacactggcctttctt	741
Db	40084	TGTCTGGCATGTGCTGGGATGATGTTGATAAACTGCTACATTTATTCACTGGCTTTTCT	40025
QY	742	tgaacaccacctcaagcaaaaaaacacacttaggtgacattattggagatgaagt	801
Db	40024	GGAACTATTGATTCAG-----AAACCATACCTTACATGACCTTATTTGGAGACAGGGT	39971
QY	802	ctttatagagatgcttaagtttaacagagacttttaaagcggctctattccatttaag	861
Db	39970	CTTTACAGAGGTAATCAAGTTAAATGAGGTCCTTAGGACAGATCCTATCCAAAT--GTG	39913
QY	862	aatggtgtccctacaaaaggaagaacttggacagagagtattgtacacttgttgtgtga	921
Db	39912	AGTGGTGTCTTTACAAAATAAANA-----AGAAAGAAATTTGGACACAGACAAGAGGG	39859
QY	922	gagacaacgtgaggagctgaagggagcagtcacaagctcagagaagggctgacccctatt	981
Db	39858	AGGACAAATGTAAGAAGATGAGATGACCATCGACGAGAGAGACGGGTATCTTCAAGC	39799
QY	982	cacacttgagcaaacagtcattgttgggttcgatatagatgagagtatcccccaagactcaca	1041
Db	39798	CAAGGAGAGAGGCTTGGACAGATCCTTCTTATTGGCCCTTAAAGAACCCCAACCTGGCA	39739

RESULT	3
CNS01DX0	
LOCUS	
DEFINITION	CNS01DX0 163317 bp DNA PRI 25-MAY-2000 Human chromosome 14 DNA sequence *** IN PROGRESS *** BAC C-238821 of library Caltech-D from chromosome 14 of Homo sapiens (Human), complete sequence.

ALL38895	
VERSION	3
GI:	7413823
HTG:	HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 163317)

AVAILABILITY	TITLE	JOURNAL	COMMENT
Genoscope.	Direct Submission		
	Submitted (23-MAY-2000)		to the EMBL/GenBank/DBJ databases
	On Apr 4, 2000		this sequence version replaced gi:7378660.
	IMPORTANT:		This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continue. The sequence may be contaminated with foreign sequence from <i>E.coli</i> , yeast, vector, phage, etc. . . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the Sp6 end.

```

Upstream BAC (overlapping the T7 end) : R-1012E15
Downstream BAC (overlapping the SP6 end) : R-930C11
-----
Overall quality chart :
range      : bases
0          : 3
1 - 9      : 532
10 - 19    : 1854
20 - 29    : 5241
30 - 39    : 13843
40 - 49    : 12446
50 - 59    : 12756
60 - 69    : 26972
70 - 79    : 47065

```

80 - 89	:	30271
90 - 99	:	12334
-----		
Percentage of bases with a quality value >= 40 : 86 %.		
FEATURES	Location/Qualifiers	
source	1..163317	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="14"	
	/clone="C-2588E21"	
	/clone_lib="CalTech-D"	
BASE COUNT	46800 a 31655 c 34205 g 50654 t 3 others	
ORIGIN		
Query Match 30.0%; Score 317.6; DB 35; Length 163317;		
Best Local Similarity 64.5%; Pred. NO. 2.8e-76;		
Matches 619; Conservative 0; Mismatches 299; Indels 42; Gaps		
Qy	107	ttacactgcacactgtgtgcagcagccagaatttaggagatgtatggacaggggtcggg 166
Db	417	TTATAAAGTCATCTTGTAATCAAAAGCCAGGATTTAGAGAGTAGGCCATTCCTCCATAGAAA 476
Qy	167	gaagaagagaaggg-taaagaaagacagcacgtttaaagtcataaacagctccaggaga 225
Db	477	GAGAGAAATGAAGAGGGTTAAGAGAAAGATAAAGAGCTATACTTCAAAAGGTCCTACAAGA 536
Qy	226	ctatctgtagaataaacatcagaccatcaggagaaattgatatcattg-----ttttt 277
Db	537	TTATTATAGAAACAGCATCAGACATGAACACTGACAAAATATTTCCTCCGATGTTCAT 596
Qy	278	caatgggatatgcacagggaactttccatctctgattataaaaaataactactgtgcactaaa 337
Db	597	GGTCTGGTATCATCAAGGGAACCTTGCACTCTGATTAAAAATTAATTATTCTGGCATTAAA 656
Qy	338	tccaattggaatgccccacacaatttatcttccacttcactgcctaccatatactctga 397
Db	657	TCAGATTGGATACACCCACACGATTATCTTCTATTTCATCTGCTGCTACCATATGCCCTGA 716
Qy	398	cgtggcggagcagaagcattccctccgcttctgtataaatagtactttgtaaatatttggga 457
Db	717	CATTGTGGTAGAGAAACATCTCTTTCATCTTGATAAGTAGTACTTTGTAAATAATTAGA 776
Qy	458	gacggagactctgg-----tgacagggaacacgtacataaacggcgctgttt 502
Db	777	GACGGGAGCTCTGGATCGCTGACAAATTCGTACAGGGGAACATGTGCAATCGAGCCCGTTA 836
Qy	503	atcatgttcctgatagagggccctctttgacgtacaggaccccccaaacagtcaggatgctg 562
Db	837	TTCATGTCTCCGAGAGAGACTCTCTTTGAGATGTGGGACCCCAAGACATTCGTGNAACGTG 896
Qy	563	tgaatttctctccatgaagccttgttccaaattagcaaccattggagggaagcaggctgcga 622
Db	897	AGAAATTCCTTTCCCAACATCTTGTTTCAGAAATTAGCAGCAATTTGGAGGAGCATGTTGGG 956
Qy	623	ctgtctaccacaagtggcactttccaagagacacacatatattggagcaagacacattttgc 682
Db	957	CACCTCI--AAATGAGCCATGCTGCCCCAAGGGGCCACATACATCAGTGAGACATTTTTCG 1013
Qy	683	tggctgaactgtgtgtgtaagctgat-aaactgctattattataaactggcctttctt 741
Db	1014	TGTCTGGCATGTGCTGGATAAGCTTGATAAACTGCTACATTTATTTCAACTGGCTTTTCTCT 1073
Qy	742	tgaacacccctctcaaggaaaaaaacacacttaggtgagattatttggagatgaagt 801
Db	1074	GGNACATTGATCTCAG-----AAACCATACCTTAACATGACCTTATTTTGGAGACAGGGT 1127
Qy	802	ctttatagatgtcttaagtttaaacgagacttttaaaggcggctctatttccatttaagt 861
Db	1128	CTTTACAGAGGTAAACAAGTTAAAAATGAGGTCCTTAGCACAGATCCTTAATCCAAT--GTG 1185
Qy	862	aatgtgtctcttacaagaagaaactgggacagaggttatgtacactgtgtgtgtgtga 921



Db 1186 AGTGGTCTCTTACAAAAA-----AGAAAGAAATTTGGACACACACAAAGAGGG 1239

Qy 922 gagacaagctgagagctgaagaggagcagctacaagtcagagaaaggtgacccctatt 981

Db 1240 AGGACAATGTGAAAGATGAAGATGACATCGACGAGAGAGAGGGCTATCTTCAAGC 1299

Qy 982 cacactgagacaaccagctatgtgtggtcgatagatgagatgaccccaagactcaca 1041

Db 1300 CAAGGAGAGAGGCTCGACACAGATCTTCTTATGGCCCTTAGAAGAACCCACCCCTGCCA 1359

RESULT 4

CNS05TBP/c 195713 bp DNA HTG 25-MAY-2000

LOCUS Homo sapiens chromosome 14 clone R-1012E15, \*\*\* SEQUENCING IN

DEFINITION PROGRESS \*\*\*, 9 ordered pieces.

VERSION AL163952.2 GI:9213040

KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 195713)

TITLE Direct Submission

JOURNAL Submitted (23-MAY-2000) to the EMBL/GenBank/DBJ databases

COMMENT On Jul 15, 2000 this sequence version replaced gi:8218003. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. . . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the SP6 end.

Upstream BAC (overlapping the T7 end) : C-2588E21

Downstream BAC (overlapping the SP6 end) : R-813120 Contigs composition :

14393 bp contig from 1 to 14393

1243 bp contig from 14494 to 15736

2086 bp contig from 15837 to 17922

1957 bp contig from 18023 to 19979

8034 bp contig from 20080 to 28113

5056 bp contig from 28214 to 33269

87497 bp contig from 33370 to 120866

68538 bp contig from 120967 to 189504

6109 bp contig from 189605 to 195713

-----

Overall quality chart :

Range : bases

0 : 1227

1 - 9 : 1625

10 - 19 : 1725

20 - 29 : 3285

30 - 39 : 13973

40 - 49 : 15406

50 - 59 : 10828

60 - 69 : 19441

70 - 79 : 47709

80 - 89 : 47803

90 - 99 : 32691

-----

Percentage of bases with a quality value >= 40 : 88 %.

\* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 14393: contig of 14393 bp in length

\* 14394 14493: gap of 100 bp

\* 14494 15736: contig of 1243 bp in length

\* 15737 15836: gap of 100 bp

\* 15837 17922: contig of 2086 bp in length

\* 17923 18022: gap of 100 bp

\* 18023 19979: contig of 1957 bp in length

\* 19980 20079: gap of 100 bp

\* 20080 28113: contig of 8034 bp in length

\* 28114 28213: gap of 100 bp

\* 28214 33269: contig of 5056 bp in length

\* 33270 33369: gap of 100 bp

\* 33370 120866: contig of 87497 bp in length

\* 120867 120966: gap of 100 bp

\* 120967 189504: contig of 68538 bp in length

\* 189505 189604: gap of 100 bp

\* 189605 195713: contig of 6109 bp in length.

Location/Qualifiers

1. 195713

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="14"

/clone\_lib="R-1012E15"

/clone="R-1012E15"

BASE COUNT 59397 a 40418 c 39000 g 55988 t 910 others

ORIGIN

Query Match 30.0%; Score 317.6; DB 87; Length 195713;

Best Local Similarity 64.5%; Pred. NO. 2.8e-76;

Matches 619; Conservative 0; Mismatches 299; Indels 42; Gaps 8;

QY 107 ttacacttcacacctgtggtcagcagcagaatttagggatgtgtagggacaggggtcggg 166

Db 39222 TTATAAGTCATCTGTGTAATCAAAAGCCAGGATTTAGAGAGTAGGCCATTCCTCCATAGAA 39163

QY 167 gaaagagggagggg-taaaggaagacagcagcagcttaaaagtcacacacagctcccgaggaga 225

Db 39162 GAGAGAAATGAAAGGGTTAAAGAAAGATAAAAGCTACTTCAAAAGGTCTCTACAAGA 39103

QY 226 ctatctgtagaataaacatcacagaccatgaggagaattgatatcattg-----ttttt 277

Db 39102 TTATTTATAGAACACACATCAGACAAATGAAACTGACAAATATTTTCCCGATGTTACAT 39043

QY 278 caatgggtatcgcaagggagactttccatctgattataaaataattactctgtggcactaaa 337

Db 39042 GGTCGTGATCATCAAGGAGAACCTTGCACTGATTAATAAATAATATTATTGCTGGCATATA 38983

QY 338 tccaattggaaatgccccacacacaaatttatcttcacacttcacgtctacacatatgcctga 397

Db 38982 TCCAGTTGGAAATCACCACACAGATTTATCTTCTATTTCATGCTGTACCATATGCTGA 38923

QY 398 cgtggcgggagcagaagcattccctccgctctgataaataagtagtactttgataatttga 457

Db 38922 CATTTGGTACAGAAACATCTCTTTCATTCGTATAGTAGTACTTTGAAATATTTAGA 38863

QY 458 gacgggagctctgg-----tgacagggaacacgtacacaccccgccctgttt 502

Db 38862 GACGGGAGCTCTGGATCGCTGACAAATCTGACAGGGAACATGTGCAATCGAGCCGCTTA 38803

QY 503 atcatgttcccgatagaggccctctttgaogtacaggaagaccccccaacagtcaggaatgcg 562

Db 38802 TTCATGTTCTCGAGAGAGACTCTCTTTGAGATGTGGGCCCAAAAGCATTCGTGAAACTG 38743

QY 563 tgaatttccttcacatgaagccttcttcacaaattagcaacattggaggagcagcgtcga 622

Db 38742 AGAATTCCTTTCCCAACATTTCTTTGATAGAAATTTAGCAGCAATTTGAGGAGCATGGTGGG 38683

QY 623 ctgtctaccacaagtgagcactttccaaagagcacacatatatttggagcagacatttgc 682

Db 38682 CACTCT---ACATGAGCCATGCTGCCCCCAAGGGCCACATACATACATGAGAGATTTGC 38626

QY 683 tggctgactggtgctgtgtaagctgat-aaactgctatatatttataaactggctttctt 741





	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
misc_feature	8107. .10506		/note="assembly_fragment"			
misc_feature	10607. .13835		/note="assembly_fragment"			
misc_feature	13936. .18202		/note="assembly_fragment"			

Query Match	5.3%	Score 56.4;	DB 53;	Length 203112;
Best Local Similarity	56.5%;	Pred. No. 0.00013;		
Matches 105;	Conservative 0;	Mismatches 81;	Indels 0;	Gaps
QY 780	tgacattattggcgtgaactctttatcagatgctttaagtttaaacagagacttttaaa	839		
Db 23087	TGTCATTATTGGAGATAGACTCTTTGTAGAGACAATTAGTTAAATGAGTCTTTGGT	23146		
QY 840	gcgcgctctattccatttaatgaatggtgctccctcaaaagggaagaactgggacagaggt	899		
Db 23147	ATGGGCTGTAATCCAAATCAAGTTGTGTCTTACAGGAGAGGAAAAATGTGGACACAAT	23206		
QY 900	aigtacattgtgtgtgtgagagacaaagctgaggagctgaaggaggagacagctacaagt	959		
Db 23207	ACAGACACAGGGAGAGATGATGTGAAGTTCACAGAAAGATGAGAAGATAGACATCTCAAAAC	23266		
QY 960	cagaga	965		
Db 23267	CAGAGA	23272		

RESULT		7	
AC012056			
LOCUS		DNA	HTG
DEFINITION	Homo sapiens chromosome 4 clone RP11-408C24 map 4,		WORKING DRAFT
	SEQUENCE, 35 unordered pieces.		
ACCESSION	AC012056		
VERSION	AC012056.3	GI-7329286	
KEYWORDS	HTG; HTGS_PHASE1;	HTGS_DRAFT.	
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 87795)  
Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,  
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,  
Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,  
DeRellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,  
Fenster, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,  
Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howard, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Landers, T., Larcocque, K., Lehotzky, J., Levine, R.,  
Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M.,  
McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,  
Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M.,  
Peterson, K., Plerre, N., Pisan, C., Pollara, V., Raymond, C.,  
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Tirrell, A.,  
Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Viel, R., Wilson, B.,  
Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wyman, D.,  
Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and  
Zody, M.

Direct Submission  
Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6970678.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WITR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L6799  
Center clone name: 456\_I.21  
-----

\* NOTE: This record contains 90 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 838: contig of 838 bp in length  
839 938: gap of 100 bp  
939 1822: contig of 884 bp in length  
1823 1922: gap of 100 bp  
1923 2805: contig of 883 bp in length  
2806 3804: contig of 899 bp in length  
3805 3904: gap of 100 bp  
3905 4736: contig of 832 bp in length  
4737 4836: gap of 100 bp  
4837 5714: contig of 878 bp in length  
5715 5814: gap of 100 bp  
5815 6680: contig of 866 bp in length  
6681 6780: gap of 100 bp  
6781 7635: contig of 855 bp in length  
7636 7735: gap of 100 bp  
7736 8633: contig of 898 bp in length  
8634 8733: gap of 100 bp  
8734 9608: contig of 875 bp in length

9609 9708: gap of 100 bp  
9709 10593: contig of 885 bp in length  
10594 10693: gap of 100 bp  
10694 11689: contig of 896 bp in length  
11690 12552: contig of 863 bp in length  
12553 12652: gap of 100 bp  
12653 13551: contig of 899 bp in length  
13552 13651: gap of 100 bp  
13652 14545: contig of 894 bp in length  
14546 14645: gap of 100 bp  
14646 15549: contig of 904 bp in length  
15550 15649: gap of 100 bp  
15650 16508: contig of 859 bp in length  
16509 16608: gap of 100 bp  
16609 17483: contig of 875 bp in length  
17484 17583: gap of 100 bp  
17584 18457: contig of 874 bp in length  
18458 18557: gap of 100 bp  
18559 19008: contig of 851 bp in length  
19009 20392: contig of 884 bp in length  
20393 20492: gap of 100 bp  
20493 21359: contig of 867 bp in length  
21360 21459: gap of 100 bp  
21460 22363: contig of 904 bp in length  
22364 22463: gap of 100 bp  
22464 23357: contig of 894 bp in length  
23358 23457: gap of 100 bp  
23458 24293: contig of 836 bp in length  
24294 24393: gap of 100 bp  
24394 25282: contig of 889 bp in length  
25283 25382: gap of 100 bp  
25383 26268: contig of 886 bp in length  
26269 26368: gap of 100 bp  
26369 27263: contig of 895 bp in length  
27274 27363: gap of 100 bp  
27364 28220: contig of 857 bp in length  
28221 28320: gap of 100 bp  
28321 29167: contig of 847 bp in length  
29168 29267: gap of 100 bp  
29268 30137: contig of 870 bp in length  
30138 30237: gap of 100 bp  
30238 31129: contig of 892 bp in length  
31130 31229: gap of 100 bp  
31230 32101: contig of 872 bp in length  
32102 32201: gap of 100 bp  
32202 33124: contig of 921 bp in length  
33123 33224: gap of 100 bp  
33223 34101: contig of 879 bp in length  
34102 34201: gap of 100 bp  
34203 35102: contig of 901 bp in length  
35103 35202: gap of 100 bp  
35203 36053: contig of 853 bp in length  
36054 36155: gap of 100 bp  
36156 37060: contig of 905 bp in length  
37061 37160: gap of 100 bp  
37161 38066: contig of 906 bp in length  
38067 38166: gap of 100 bp  
38167 39065: contig of 899 bp in length  
39066 39165: gap of 100 bp  
39166 40029: contig of 864 bp in length  
40030 40129: gap of 100 bp  
40130 40977: contig of 848 bp in length  
40978 41077: gap of 100 bp  
41078 41947: contig of 870 bp in length  
41948 42047: gap of 100 bp  
42049 42942: contig of 895 bp in length  
42943 43042: gap of 100 bp  
43043 43909: contig of 867 bp in length  
43910 44009: gap of 100 bp  
44010 44905: contig of 896 bp in length  
44906 45005: gap of 100 bp

TITLE  
JOURNAL  
COMMENT



```
* 7190 8621: contig of 1432 bp in length
* 8622 8721: gap of 100 bp
* 8722 9817: contig of 1096 bp in length
* 9818 9917: gap of 100 bp
* 9918 10339: contig of 1022 bp in length
* 10340 11039: gap of 100 bp
* 11040 12650: contig of 1611 bp in length
* 12651 12750: gap of 100 bp
* 12751 14251: contig of 1501 bp in length
* 14252 14351: gap of 100 bp
* 14352 16578: contig of 2227 bp in length
* 16579 18045: contig of 1367 bp in length
* 18046 18145: gap of 100 bp
* 18146 19697: contig of 1552 bp in length
* 19698 19797: gap of 100 bp
* 19798 21993: contig of 2196 bp in length
* 21994 22093: gap of 100 bp
* 22094 23427: contig of 1334 bp in length
* 23428 23527: gap of 100 bp
* 23528 26364: contig of 2837 bp in length
* 26365 26464: gap of 100 bp
* 26465 28668: contig of 2204 bp in length
* 28669 28768: gap of 100 bp
* 28769 30370: contig of 1602 bp in length
* 30371 30470: gap of 100 bp
* 30471 33691: contig of 3221 bp in length
* 33692 33791: gap of 100 bp
* 33792 35652: contig of 1861 bp in length
* 35653 35752: gap of 100 bp
* 35753 37693: contig of 1943 bp in length
* 37694 37793: gap of 100 bp
* 37794 40518: contig of 2723 bp in length
* 40519 40618: gap of 100 bp
* 40619 43583: contig of 2965 bp in length
* 43584 43683: gap of 100 bp
* 43684 46732: contig of 3049 bp in length
* 46733 46832: gap of 100 bp
* 46833 48822: contig of 1990 bp in length
* 48823 48922: gap of 100 bp
* 48923 49655: contig of 733 bp in length
* 49656 49755: gap of 100 bp
* 49756 54236: contig of 4481 bp in length
* 54237 54336: gap of 100 bp
* 54337 57763: contig of 3427 bp in length
* 57764 57863: gap of 100 bp
* 57864 61081: contig of 3218 bp in length
* 61082 61181: gap of 100 bp
* 61182 64263: contig of 3082 bp in length
* 64264 64363: gap of 100 bp
* 64364 69021: contig of 4658 bp in length
* 69022 69121: gap of 100 bp
* 69122 73024: contig of 3903 bp in length
* 73025 73124: gap of 100 bp
* 73125 77636: contig of 4512 bp in length
* 77637 77736: gap of 100 bp
* 77737 82200: contig of 4464 bp in length
* 82201 82300: gap of 100 bp
* 82301 87681: contig of 5381 bp in length
* 87682 87781: gap of 100 bp
* 87782 94716: contig of 6935 bp in length
* 94717 94816: gap of 100 bp
* 94817 100108: contig of 5292 bp in length
* 100109 100208: gap of 100 bp
* 100209 105102: contig of 4894 bp in length
* 105103 105202: gap of 100 bp
* 105203 110809: contig of 5607 bp in length
* 110810 110909: gap of 100 bp
* 110910 116649: contig of 5740 bp in length
* 116650 116749: gap of 100 bp
* 116750 122701: contig of 5952 bp in length
* 122702 122801: gap of 100 bp
* 122802 127928: contig of 5127 bp in length
```

```
* 127929 128028: gap of 100 bp
* 128029 135192: contig of 7164 bp in length
* 135193 135292: gap of 100 bp
* 135293 141464: contig of 6172 bp in length
* 141465 141564: gap of 100 bp
* 141565 148339: contig of 6775 bp in length
* 148340 148439: gap of 100 bp
* 148440 155497: contig of 7038 bp in length
* 155498 155597: gap of 100 bp
* 155598 163096: contig of 7499 bp in length
* 163097 163196: gap of 100 bp
* 163197 170875: contig of 7679 bp in length
* 170876 170975: gap of 100 bp
* 170976 179783: contig of 8808 bp in length
* 179784 179883: gap of 100 bp
* 179884 190675: contig of 10792 bp in length
* 190676 190775: gap of 100 bp
* 190776 198653: contig of 7878 bp in length
* 198654 198753: gap of 100 bp
* 198754 207949: contig of 9196 bp in length
* 207950 208049: gap of 100 bp
* 208050 219082: contig of 11033 bp in length
* 219083 219182: gap of 100 bp
* 219183 232806: contig of 13624 bp in length
* 232807 232906: gap of 100 bp
* 232907 246467: contig of 13561 bp in length
* 246468 246567: gap of 100 bp
* 246568 259249: contig of 12682 bp in length
* 259250 259349: gap of 100 bp
* 259350 271902: contig of 12553 bp in length
* 271903 272002: gap of 100 bp
* 272003 288943: contig of 16941 bp in length
* 288944 289043: gap of 100 bp
* 289044 302367: contig of 13324 bp in length
* 302368 302467: gap of 100 bp
* 302468 323907: contig of 21440 bp in length
* 323908 324007: gap of 100 bp
* 324008 351436: contig of 27429 bp in length.
```

## FEATURES

```
source
1..351436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-2H15"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1..1100
/note="assembly_fragment"
```

```
Query Match 5.2%; Score 54.6; DB 36; Length 110000;
Best Local Similarity 57.6%; Pred. No. 0.00038;
Matches 117; Conservative 0; Mismatches 84; Indels 2; Gaps 1;
```

```
QY 776 aggggtgacattattggagatgaagtctttatagatgcttaagttaaacgagacttt 835
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88169 AGTGGACCTTTTGGAAATACAGGCTTCACAGGTAATCAAGTTAAACGAAGTCAT 88110

QY 836 taaagccggc:ctattccatttaataatgaatggtgctccctacaaaggaagaactgggacag 895
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88109 TAGAGTGAGCCTTAATCCACATGACGTGGTCTCTTATATAAAGAGAGAAATTTGGACAG 88050

QY 896 aggtatgtacactgtgtgtgtgagagacacactgagagctgaagagagcagctac 955
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88049 AG--AGGCAGACATGAGGGATGACAATGTGAGGAGACACAGGGAGAGATGACCATCTAC 87992

QY 956 aagtcagaganaagctgacctt 978
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87991 AAGCCAGGAATGCTGACGTTT 87969
```

```
RESULT 10
AC008072/c
LOCUS
DEFINITION Homo sap:ens clone RP11-408N22, complete sequence.
ACCESSION AC008072
PRI
16-DEC-1999
```



```

VERSION AC008072.3 GI:6587924
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 206177)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 206177)
TITLE Waterston, R.H.
AUTHORS Direct Submission
JOURNAL Submitted (17-JUL-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 206177)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Dec 16, 1999 this sequence version replaced gi:5732152.
Center project name: NH0408N22.
FEATURES
source
1. 206177
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-408N22"
BASE COUNT 58662 a 41215 c 44233 g 62067 t
ORIGIN
Query Match 5.2%; Score 54.6; DB 10; Length 206177;
Best Local Similarity 68.7%; Pred. No. 0.0004;
Matches 90; Conservative 0; Mismatches 39; Indels 2; Gaps 1;
QY 776 aggtgacattattgagatgaagtctttatagatgcttaagttaacagagacttt 835
Db 180501 AGTGTGACCTATTGGAGATAGATGCTTTACAGAGGTATCAAGTTAAATGAAGTTAT 180442
QY 836 taaagccggctctatccatttaataatgaatggtgtccctacaaggaagaactgggacag 895
Db 180441 TAGAGTGGGCCCTAATCCGAT--ATGACTGGTGTCTCTATAGAGAGAAAATGTGGACAC 180384
QY 896 aggtatgtatca 906
Db 180383 AGAGAGGCAGCA 180373
RESULT 11
CNS01RHA 167095 bp DNA PRI 25-MAY-2000
LOCUS Human chromosome 14 DNA sequence *** IN PROGRESS *** BAC R-89K22 of
DEFINITION library RPCI-11 from chromosome 14 of Homo sapiens (Human),
complete sequence.
ACCESSION AL161666
VERSION AL161666.2 GI:7406477
KEYWORDS HTG; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167095)
Genoscope.
Direct Submission
Submitted (23-MAY-2000) to the EMBL/GenBank/DBJ databases
On Apr 3, 2000 this sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc. . . even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-893D14
Downstream BAC (overlapping the SP6 end) : R-769B21
Overall quality chart :
Range : bases
0 - 9 : 495
1 - 9 : 1451
10 - 19 : 4890
20 - 29 : 19367
30 - 39 : 12338
40 - 49 : 12811
50 - 59 : 27547
60 - 69 : 48484
70 - 79 : 28718
80 - 89 : 10993
90 - 99 :
Percentage of bases with a quality value >= 40 : 84 %.
FEATURES
source
1. 167095
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="RPCI-11"
/clone="R-89K22"
BASE COUNT 51844 a 32959 c 32634 g 49657 t 1 others
ORIGIN
Query Match 5.1%; Score 53.8; DB 35; Length 167095;
Best Local Similarity 61.4%; Pred. No. 0.00065;
Matches 105; Conservative 0; Mismatches 62; Indels 4; Gaps 1;
QY 779 gtgacattattgagatgaagtctttatagatgcttaagttaacagagacttttaa 838
Db 118767 GTGACATTTTATAGATAGAGTTTTCACAGAGGTATTAAGTTAAATGATCATAG 118826
QY 839 agccggctctatccatttaataatgaatggtgtccctacaaggaagaactgggacag 898
Db 118827 GGTGGCTCTACCAAT---ATGACTGGTGTCTTATAAAGAGGAAATTTGGACACAGA 118882
QY 899 tatgtacactgtgtgtgtgagagacacgtgagagctgagagctgagagagc 949
Db 118883 GATATGAATAGGCAGACATATGGGAGACCCAGGAGAGAGAAACATC 118933
RESULT 12
CNS01RIB 181905 bp DNA HTG 08-AUG-2000
LOCUS Homo sapiens chromosome 14 clone R-769B21, *** SEQUENCING IN
DEFINITION PROGRESS***, in ordered pieces.
ACCESSION AL163052
VERSION AL163052.3 GI:9756024
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 181905)
Direct Submission
Submitted (08-AUG-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
On Aug 9, 2000 this sequence version replaced gi:9213023.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,

```

phage, etc. . . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the SP6 end.

Upstream BAC (overlapping the T7 end) : C-2149C7

Downstream BAC (overlapping the SP6 end) : R-89K22

-----  
Overall quality chart :  
Range : bases

```

0 - 9 : 14
10 - 19 : 51
20 - 29 : 125
30 - 39 : 568
40 - 49 : 4404
50 - 59 : 6346
60 - 69 : 4584
70 - 79 : 11431
80 - 89 : 44965
90 - 99 : 109417

```

Percentage of bases with a quality value >= 40 : 99 %.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

#### FEATURES

source Location/Qualifiers

```

1..181905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="RPCI-11"
/clone="R-769B21"

```

BASE COUNT 57549 a 35621 c 34265 g 54470 t  
ORIGIN

```

Query Match      5.1% Score 53.8; DB 87; Length 181905;
Best Local Similarity 61.4%; Pred. No. 0.00065;
Matches 105; Conservative 0; Mismatches 62; Indels 4; Gaps 1;

QY 779 gtgcattatttgagatgaagcttttatagagatgcttaagttaaacagagacttttaa 838
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173586 GTGCATTTTATAGATAGATAGGTTTACAGAGGTAATTAAATGATCATTTAG 173527

QY 839 agccgctctattccatttaataatgcttccctcaagagaaactggagacagag 898
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173526 GGTGGGCTCTACCAAT----ATGACTGGTGCTCTATAAAGAGAAATTTGGACACAGA 173471

QY 899 tatgtacacttgctgtgtgtgagagacacgtgagagctgagagggagc 949
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173470 GATATGAATAGAGCAAGACATGGGAAGACCCAGGGAGAGAAACATC 173420

```

RESULT 13  
AC067885/c  
LOCUS  
DEFINITION Homo sapiens chromosome 10 clone RP11-276J19 map 10, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 81 unordered pieces.  
AC067885  
VERSION  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE  
human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 209706)  
Blair, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE  
Homo sapiens chromosome 10, clone RP11-276J19  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 209706)  
Blair, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,  
Boguslavsky, L., Boukhgaiter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collins, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
Meidrim, J., Meneus, D., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Olivier, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

#### Direct Submission

Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 3: 2000 This sequence version replaced gi:7651988.

All repeats were identified using RepeatMasker:

Smit, A. F. A. & Green, P. (1996-1997)

http://fpc.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8017

Center clone name: 276\_JJ\_19

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 81 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

* 1 1001: contig of 1001 bp in length
* 1002 1101: gap of 100 bp
* 1102 2321: contig of 1220 bp in length
* 2322 2421: gap of 100 bp
* 2422 3702: contig of 1281 bp in length
* 3703 3802: gap of 100 bp
* 3803 4852: contig of 1050 bp in length
* 4853 4952: gap of 100 bp
* 4953 6025: contig of 1073 bp in length
* 6026 6125: gap of 100 bp
* 6126 7231: contig of 1106 bp in length
* 7232 7331: gap of 100 bp
* 7332 8642: contig of 1311 bp in length
* 8643 8742: gap of 100 bp
* 8743 10154: contig of 1412 bp in length
* 10155 10254: gap of 100 bp
* 10255 11858: contig of 1604 bp in length
* 11859 11958: gap of 100 bp
* 11959 13442: contig of 1484 bp in length
* 13443 13542: gap of 100 bp
* 13543 14674: contig of 1132 bp in length
* 14675 14774: gap of 100 bp
* 14775 15959: contig of 1185 bp in length
* 15960 16059: gap of 100 bp
* 16060 17503: contig of 1444 bp in length
* 17504 17603: gap of 100 bp
* 17604 19298: contig of 1695 bp in length
* 19299 19398: gap of 100 bp
* 19399 20995: contig of 1597 bp in length
* 20996 21095: gap of 100 bp
* 21096 22482: contig of 1387 bp in length

```

Mon Nov 6 10:14:12 2000

\* 22483 22582: gap of 100 bp  
\* 22583 24596: contig of 2014 bp in length  
\* 24597 24696: gap of 100 bp  
\* 24697 26841: contig of 2145 bp in length  
\* 26842 26941: gap of 100 bp  
\* 26942 27988: contig of 1047 bp in length  
\* 27989 28088: gap of 100 bp  
\* 28089 29224: contig of 1136 bp in length  
\* 29225 29324: gap of 100 bp  
\* 29325 30857: contig of 1533 bp in length  
\* 30858 30937: gap of 100 bp  
\* 30938 32690: contig of 1733 bp in length  
\* 32691 32790: gap of 100 bp  
\* 32791 34377: contig of 1587 bp in length  
\* 34378 34477: gap of 100 bp  
\* 34478 35884: contig of 1407 bp in length  
\* 35885 35984: gap of 100 bp  
\* 35985 37699: contig of 1715 bp in length  
\* 37700 37799: gap of 100 bp  
\* 37800 39419: contig of 1620 bp in length  
\* 39420 39519: gap of 100 bp  
\* 39520 41137: contig of 1618 bp in length  
\* 41138 41237: gap of 100 bp  
\* 41238 42418: contig of 1179 bp in length  
\* 42417 42516: gap of 100 bp  
\* 42517 43560: contig of 1044 bp in length  
\* 43561 43660: gap of 100 bp  
\* 43661 45821: contig of 2161 bp in length  
\* 45822 45921: gap of 100 bp  
\* 45922 47891: contig of 1970 bp in length  
\* 47892 47991: gap of 100 bp  
\* 47992 50142: contig of 2151 bp in length  
\* 50143 50242: gap of 100 bp  
\* 50243 51935: contig of 1693 bp in length  
\* 51936 52035: gap of 100 bp  
\* 52036 53838: contig of 1803 bp in length  
\* 53839 53938: gap of 100 bp  
\* 53939 55920: contig of 1982 bp in length  
\* 55921 56020: gap of 100 bp  
\* 56021 58049: contig of 2029 bp in length  
\* 58050 58149: gap of 100 bp  
\* 58150 59795: contig of 1646 bp in length  
\* 59796 59895: gap of 100 bp  
\* 59896 61145: contig of 1250 bp in length  
\* 61146 61245: gap of 100 bp  
\* 61246 63280: contig of 2035 bp in length  
\* 63281 63380: gap of 100 bp  
\* 63381 64917: contig of 1537 bp in length  
\* 64918 65017: gap of 100 bp  
\* 65018 66772: contig of 1755 bp in length  
\* 66773 66872: gap of 100 bp  
\* 66873 68393: contig of 1521 bp in length  
\* 68394 68493: gap of 100 bp  
\* 68494 69911: contig of 1418 bp in length  
\* 69912 70011: gap of 100 bp  
\* 70012 72745: contig of 2734 bp in length  
\* 72746 72845: gap of 100 bp  
\* 72846 75038: contig of 2193 bp in length  
\* 75039 75138: gap of 100 bp  
\* 75139 77737: contig of 2599 bp in length  
\* 77738 77837: gap of 100 bp  
\* 77838 80616: contig of 2779 bp in length  
\* 80617 80716: gap of 100 bp  
\* 80717 82403: contig of 1687 bp in length  
\* 82404 82503: gap of 100 bp  
\* 82504 85589: contig of 3086 bp in length  
\* 85590 85689: gap of 100 bp  
\* 85690 88175: contig of 2486 bp in length  
\* 88176 88275: gap of 100 bp  
\* 88276 90565: contig of 2290 bp in length  
\* 90566 90665: gap of 100 bp  
\* 90666 93033: contig of 2368 bp in length  
\* 93034 93133: gap of 100 bp

\* 93134 96273: contig of 3140 bp in length  
\* 96274 96373: gap of 100 bp  
\* 96374 99128: contig of 2755 bp in length  
\* 99129 99228: gap of 100 bp  
\* 99229 101437: contig of 2209 bp in length  
\* 101438 101537: gap of 100 bp  
\* 101538 104158: contig of 2621 bp in length  
\* 104159 104258: gap of 100 bp  
\* 104259 107778: contig of 3520 bp in length  
\* 107779 107878: gap of 100 bp  
\* 107879 110744: contig of 2866 bp in length  
\* 110745 110844: gap of 100 bp  
\* 110845 112903: contig of 2059 bp in length  
\* 112904 113003: gap of 100 bp  
\* 113004 116853: contig of 3850 bp in length  
\* 116854 116953: gap of 100 bp  
\* 116954 120359: contig of 3406 bp in length  
\* 120360 120459: gap of 100 bp  
\* 120460 123407: contig of 2948 bp in length  
\* 123408 123507: gap of 100 bp  
\* 123508 127440: contig of 3933 bp in length  
\* 127441 127540: gap of 100 bp  
\* 127541 131078: contig of 3538 bp in length  
\* 131079 131178: gap of 100 bp  
\* 131179 134271: contig of 3093 bp in length  
\* 134272 134371: gap of 100 bp  
\* 134372 138409: contig of 4038 bp in length  
\* 138410 138509: gap of 100 bp  
\* 138510 142150: contig of 3641 bp in length  
\* 142151 142250: gap of 100 bp  
\* 142251 146213: contig of 3963 bp in length  
\* 146214 146313: gap of 100 bp  
\* 146314 150814: contig of 4501 bp in length  
\* 150815 150914: gap of 100 bp  
\* 150915 154889: contig of 3975 bp in length  
\* 154890 154989: gap of 100 bp

Query Match 5.1%; Score 53.8; DB 55; Length 209706;  
Best Local Similarity 67.4%; Pred. No. 0.00066;  
Matches 91; Conservative 0; Mismatches 42; Indels 2; Gaps 1;

QY 772 acctgggtgacattatttgagatgaagtccttatagagatgcttaagtttaacagaga 831  
Db 95714 ACCTCAGTGTGCTGGGTTGGAGTGGGGTCTTTACAGAGATAATTAGATAAATGAGG 95655  
QY 832 ctttaaaagcggctctattccatttaagtggtgctccctacaaggaagaactggg 891  
Db 95654 TCTTTAGGTGGACCCCTACTCCATC--TGCATGGTGTCTTATAAAGCAGACGTTAGG 95597  
QY 892 acagagggtatgtaca 906  
Db 95596 ACACAGACATGCACA 95582

## RESULT 14

AC069437  
LOCUS Homo sapiens chromosome 12 clone RP11-473C19, WORKING DRAFT  
DEFINITION AC069437.4 GI:9438750  
ACCESSION AC069437  
VERSION AC069437.4  
KEYWORDS HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

1 (bases 1 to 184511)  
Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,  
Bodota, E., Bock, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,  
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,  
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,  
Dugan-Rocha, S., Durbin, K. J., Fernandez, C., Ferraguto, D.,  
Forcum-Ransey, J., Frantz, P., Ganesh, R., Gorrell, J. H., Gorrell, L. L.,



variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

IMPORTANT: This sequence is not the entire insert of clone 77N19. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 426F10 is at 56705 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>

77N19 is from the library RPC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PCYAC2>.

# FEATURES source

1. 56804  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/Chromosome="1"  
/map="p36.2-36.3"  
/clone="RPI-77N19"  
/clone\_lib="RPC1-1"

repeat\_region 1. .245  
/note="AluSq repeat: matches 59. .303 of consensus"  
unsure complement(148. .195)  
repeat\_region /note="MLTIE repeat: matches 514. .282 of consensus"  
repeat\_region complement(1691. .1803)  
repeat\_region /note="MLTIE repeat: matches 158. .46 of consensus"  
repeat\_region complement(1895. .2070)  
repeat\_region /note="MIR repeat: matches 262. .80 of consensus"  
repeat\_region 2327. .2463  
/note="MIR2 repeat: matches 1. .146 of consensus"  
repeat\_region complement(2540. .2850)  
repeat\_region /note="MLTIC repeat: matches 466. .141 of consensus"  
repeat\_region complement(2837. .2884)  
repeat\_region /note="MLTIG repeat: matches 388. .341 of consensus"  
repeat\_region complement(2916. .3396)  
repeat\_region /note="MLTICB repeat: matches 460. .1 of consensus"  
repeat\_region complement(3395. .3533)  
repeat\_region /note="MLTIC repeat: matches 163. .6 of consensus"  
repeat\_region 4877. .4954  
/note="MIR repeat: matches 79. .155 of consensus"  
repeat\_region complement(5029. .5152)  
repeat\_region /note="MIR repeat: matches 155. .28 of consensus"  
repeat\_region 5736. .6005  
/note="6 copies 45 mer 80% conserved"  
repeat\_region 5739. .5990  
/note="7 copies 36 mer 83% conserved"  
repeat\_region 6156. .6336  
/note="MIR repeat: matches 36. .225 of consensus"  
repeat\_region complement(6393. .6547)  
repeat\_region /note="MERSA repeat: matches 164. .5 of consensus"  
repeat\_region complement(7756. .7788)  
repeat\_region /note="MIR repeat: matches 143. .111 of consensus"  
repeat\_region complement(9279. .9401)  
repeat\_region /note="MLTIB repeat: matches 390. .264 of consensus"  
repeat\_region complement(10873. .10957)  
repeat\_region /note="MIR2 repeat: matches 146. .66 of consensus"  
repeat\_region 11365. .11392  
/note="14 copies 2 mer ag 89% conserved"  
repeat\_region 11559. .11859  
/note="AluSx repeat: matches 1. .302 of consensus"  
unsure complement(11821. .11862)  
repeat\_region complement(11960. .11999)  
/note="MIR repeat: matches 150. .111 of consensus"

repeat\_region complement(13693. .13783)  
/note="MIR repeat: matches 203. .109 of consensus"  
repeat\_region complement(15229. .15267)  
/note="MIR repeat: matches 149. .111 of consensus"  
repeat\_region complement(16678. .17027)  
/note="MSTC repeat: matches 405. .19 of consensus"  
repeat\_region complement(16678. .16903)  
/note="MSTD repeat: matches 394. .166 of consensus"  
repeat\_region complement(17487. .17611)  
/note="MIR2 repeat: matches 146. .21 of consensus"  
repeat\_region 18092. .18213  
/note="MER20 repeat: matches 5. .131 of consensus"  
repeat\_region complement(18610. .18954)  
repeat\_region /note="MLTIB repeat: matches 374. .1 of consensus"  
repeat\_region 20694. .20796  
/note="MIR repeat: matches 68. .172 of consensus"  
repeat\_region 22858. .23006  
/note="MIR repeat: matches 110. .262 of consensus"  
repeat\_region complement(23392. .23542)  
/note="MERSB repeat: matches 143. .2 of consensus"  
repeat\_region complement(24370. .24444)  
/note="MIR repeat: matches 137. .63 of consensus"  
repeat\_region complement(24638. .24792)  
/note="MER20 repeat: matches 170. .6 of consensus"  
repeat\_region 25075. .25283  
/note="11 copies 19 mer 84% conserved"  
repeat\_region complement(27234. .28125)  
/note="L1PA2 repeat: matches 893. .1 of consensus"  
repeat\_region complement(27976. .28243)  
/note="L1 repeat: matches 5390. .5124 of consensus"  
repeat\_region 28243. .28444  
/note="L1 repeat: matches 4918. .5118 of consensus"  
repeat\_region 28464. .28497  
/note="17 copies 2 mer tt 100% conserved"  
repeat\_region 28509. .29116  
/note="L1MC3 repeat: matches 1850. .2486 of consensus"  
repeat\_region complement(30346. .31023)  
/note="L1MB7 repeat: matches 913. .192 of consensus"  
repeat\_region complement(31035. .31333)  
/note="AluSc repeat: matches 299. .1 of consensus"  
unsure complement(34028. .34044)  
prim\_transcript 35287. .35705  
/note="match: multiple ESTs; match: AA452856 U21460"  
repeat\_region 35591. .35623  
/note="11 copies 3 mer cca 91% conserved"  
repeat\_region 37288. .37332  
/note="3 copies 15 mer 93% conserved"  
repeat\_region 38342. .38456  
/note="MIR repeat: matches 23. .141 of consensus"  
repeat\_region 38472. .38767  
/note="AluSp repeat: matches 1. .298 of consensus"  
repeat\_region 39344. .39546  
/note="MIR repeat: matches 36. .262 of consensus"  
repeat\_region 39555. .39773  
/note="73 copies 3 mer gat 81% conserved"  
repeat\_region 39557. .39772  
/note="6 copies 36 mer 84% conserved"  
repeat\_region 39779. .39886  
/note="3 copies 36 mer 82% conserved"  
repeat\_region complement(41552. .41848)  
/note="AluSc repeat: matches 299. .1 of consensus"  
repeat\_region 44102. .44191  
/note="2 copies 45 mer 90% conserved"  
repeat\_region complement(45576. .45889)  
/note="MLTIC repeat: matches 466. .114 of consensus"  
repeat\_region complement(46697. .49070)  
/note="MLTIB repeat: matches 390. .1 of consensus"  
repeat\_region 49204. .49311  
/note="3 copies 36 mer 97% conserved"  
repeat\_region complement(49695. .49787)  
/note="MERSA repeat: matches 160. .64 of consensus"  
repeat\_region 51077. .51218  
/note="MLTIE repeat: matches 428. .563 of consensus"



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:14:58 ; Search time 254.92 seconds  
(without alignments)  
1559.121 Million cell updates/sec

Title: US-09-101-423A-2

Perfect score: 1058

Sequence: 1 attctgtgagcctattagc.....acacattcgaacgttggtc 1058

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

N\_Geneseq\_36:\*

- 1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT:\*
- 2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT:\*
- 3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT:\*
- 4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT:\*
- 5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT:\*
- 6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT:\*
- 7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT:\*
- 8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT:\*
- 9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT:\*
- 10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT:\*
- 11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT:\*
- 12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT:\*
- 13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT:\*
- 14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT:\*
- 15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT:\*
- 16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT:\*
- 17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT:\*
- 18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:\*
- 19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:\*
- 20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:\*
- 21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1058	100.0	1058	18 T72786	Metastasis inducin
2	39	3.7	437	17 T32272	Probe for the dete
3	39	3.7	437	18 T71682	Cancer detection t
4	36.8	3.5	43069	21 Z36335	Genomic sequence o
5	35	3.3	133894	17 T13635	ACNPV genomic DNA
6	34.8	3.3	1328	16 T22313	Human gene signatu
7	34.6	3.3	328	19 V61346	Forward DNA sequen
8	34.6	3.3	328	19 V58682	Prostate tumour sp
9	34.6	3.3	328	21 A06445	Human immunogenic
10	34.6	3.3	742	20 Z15036	Human gene express
11	34	3.2	704	20 Z00436	Human secreted pro
12	34	3.2	1906	21 A26302	Human secreted pro

c 13	33.6	3.2	388	16	T24148	Human gene signatu
c 14	33.4	3.2	762	21	A01861	Human colon cancer
c 15	33.2	3.1	2030	21	A34785	Human adenosine re
c 16	33	3.1	355	14	Q39730	Expressed Sequence
c 17	33	3.1	355	14	Q59142	Human brain expres
c 18	33	3.1	5763	18	V74816	Staphylococcus aur
c 19	32.6	3.1	177	13	Q29030	Low frequency repe
c 20	32.4	3.1	1236	20	Z17055	Human gene express
c 21	31.8	3.0	444	19	V61348	Forward DNA sequen
c 22	31.8	3.0	444	19	V58684	Prostate tumour sp
c 23	31.8	3.0	444	21	A06447	Human immunogenic
c 24	31.8	3.0	756	21	A01860	Human colon cancer
c 25	31.8	3.0	1817	19	V60815	Mouse ULIP-1 codin
c 26	31.8	3.0	2026	19	V46313	Human secreted pro
c 27	31.8	3.0	1664976	19	V21209	Methanococcus jann
c 28	31.6	3.0	1609	20	X02816	Pea Glucose-6-phos
c 29	31.4	3.0	300	20	Z14327	Human gene express
c 30	31.2	2.9	2513	12	Q14531	T cell protein 19.
c 31	31.2	2.9	2513	15	Q63616	T-cell transembra
c 32	31.2	2.9	3659	7	N60204	Interferon-pseudo-
c 33	31	2.9	910715	20	X20248	Borrelia burgdorfe
c 34	30.8	2.9	495	20	V87133	EST clone BD522.
c 35	30.8	2.9	1021	21	A26441	Human secreted pro
c 36	30.8	2.9	1383	21	D00098	Murine zsig49 dege
c 37	30.6	2.9	1017	19	V34170	Human secreted pro
c 38	30.6	2.9	1027	21	Z64996	Membrane-bound pro
c 39	30.6	2.9	2492	19	V34236	Human secreted pro
c 40	30.6	2.9	4732	20	X12999	Enterococcus faeca
c 41	30.6	2.9	10732	21	A10594	Gene encoding a su
c 42	30.6	2.9	29555	18	V74517	Staphylococcus aur
c 43	30.4	2.9	1002	21	A05478	Streptococcus pneu
c 44	30.4	2.9	1043	19	T98601	DNA encoding a S
c 45	30.4	2.9	2415	21	Z39572	S. pneumoniae pria

#### ALIGNMENTS

#### RESULT 1

T72786  
ID T72786 standard; DNA; 1058 BP.

XX T72786;

XX AC T72786;

XX DT 22-SEP-1997 (first entry)

XX DE Metastasis inducing DNA C5.

XX KW Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;  
osteopontin; ss.

XX OS Homo sapiens.

XX PN WO9725443-A1.

XX PD 17-JUL-1997.

XX PF 10-JAN-1997; 97WO-GB000074.

XX PR 10-JAN-1996; 96GB-0000470.

XX PA (UYLI-) UNIV LIVERPOOL.

XX PI Barracough BR, Rudland PS;

XX DR WPI; 1997-372878/34.

XX PT New isolated metastasis-inducing DNA - used to develop products to  
identify and treat patients at risk from metastatic tumours

XX PS Claim 9; Page 25; 38pp; English.

XX CC Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12

0v 820 stttaaa

820 atttaaacgacactttrtaaacggcctctattccatttaatgaatggtatc 87







RESULT 5  
T13635/C  
ID T13635 standard; cDNA; 133894 BP.  
XX AC  
XX T13635;  
DT 03-SEP-1996 (first entry)  
XX AC  
XX ACNPV genomic DNA clone 6.  
DE Autographa californica nuclear polyhedrosis virus clone 6;  
XX Autographa californica nuclear polyhedrosis virus clone 6;  
KW disruption; non-essential gene; heterologous protein production;  
KW expression vector; baculovirus; ss.  
XX OS  
XX Autographa californica nuclear polyhedrosis virus clone 6.  
XX PN  
XX WO9601320-A2.  
PD 18-JAN-1996.  
XX PF  
XX 30-JUN-1995; 95WO-IB00578.  
XX PR  
XX 04-JUL-1994; 94GB-0013420.  
XX PA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.  
XX PI  
XX Ayres M, Bishop D, Possee R;  
XX WPI; 1996-087670/09.  
DR GENBANK; L22858.  
XX  
XX Autographa californica nuclear polyhedrosis virus complete genome  
PT sequence - useful in the prodn. of vectors for enhanced  
PT heterologous protein expression, such as interleukin(s),  
PT interferon(s) and neurotoxin(s)  
XX  
XX Disclosure; Page 90-186; 122pp; English.  
XX  
XX The complete nucleotide sequence of the genome of clone 6 of the  
XX baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV)  
XX has been determined. The sequence is taken from the Genbank record  
XX L22858. The patent specification claims a polynucleotide selected from  
XX open reading frames (ORFs) 13, 20, 22-26, 28-30, 32, 38, 41-46, 50-60,  
XX 62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,  
XX 140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by  
XX the patentees. See T13636-731. Expression vectors contg. the complete  
XX genomic sequence of AcNPV, with the exception that at least one non-  
XX essential ORF is disrupted or replaced are useful for the synthesis of  
XX heterologous proteins.  
XX  
XX Sequence 133894 BP; 39195 A; 27151 C; 27347 G; 40201 T; 0 other;  
SQ

Query Match 3.3%; Score 35; DB 17; Length 133894;  
Best Local Similarity 55.3%; Pred. No. 4.9;  
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 756 aaggaagaaacacacacttaggtgacattatttgagatgaagtcctttatagagatgc 815  
DB 58352 AAACACTAATAAATCTATTCAAGAACAAATATAAGGGGTCTCAGCAGATCAATGAGGTGT 58293  
QY 816 ttaagtttaacagacttttaagcgcgtctctatccatttaagtgaatgggtgccttac 875  
DB 58292 TTAATATTTAAAGACAAATGAANAACCAATCTTTACCATTTACAGTATATAAGTGTCAAC 58233  
QY 876 aaa 878  
DB 58232 AAA 58230  
RESULT 6  
T22313/C

T22313 standard; cDNA to mRNA; 138 BP.  
T22313;  
13-SEP-1996 (first entry)  
Human gene signature HUMGS03886.  
Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
human; cloning; mapping; non-biased library; diagnosis; detection;  
cell typing; abnormal cell function; ss.  
Homo sapiens.  
WO9514772-A1.  
01-JUN-1995.  
11-NOV-1994; 94WO-JP01916.  
12-NOV-1993; 93JP-0355504.  
(MATS/) MATSUBARA K.  
(OKUB/) OKUBO K.  
Matsubara K, Ckubo K;  
WPI; 1995-206931/27.  
Identifying gene signatures in 3'-directed human cDNA library - e.g.  
for diagnosis of abnormal cell function, by preparing cDNA that  
reflects relative abundance of corresp. mRNA in specific human  
tissues  
Claim 1; Page 1097; 2245pp; Japanese.  
A single-stranded DNA (or its complementary strand or the corresp.  
double-stranded DNA) which comprises one of the 7837 "GS" sequences  
given in T19001-T26837 and which is able to hybridise to part of  
human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
sequences were obtained from 3'-directed cDNA libraries prepared  
from various human tissues: synthesis of cDNA was initiated from the  
3'-end of mRNA by using poly(rI) as the sole primer. Since the 3'-  
untranslated sequence is unique to a particular mRNA species, almost  
all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
is constructed so as to reflect accurately the relative abundance of  
different mRNAs in the particular tissue from which it was derived.  
The appearance frequency of a given GS in a cDNA library can be  
determined (esp. using primers and probes derived from the GS  
sequences) as a means of diagnosing abnormal cell function or for  
recognising different cell types.  
Sequence 138 BP; 44 A; 28 C; 19 G; 46 T; 1 other;  
SQ

Query Match 3.3%; Score 34.8; DB 16; Length 138;  
Best Local Similarity 64.6%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 51; Conservative 0; Mismatches 28;  
QY 766 aaacacacttaggtgacattatttgagatgaagtcctttatagagatgcttaagtttaa 825  
DB 134 AATACCTCATAGGTCACCTTTATTGGAAATACATTTTTTACAGAGGTAATCAAGTTAAA 75  
QY 826 acgagacttttaaagccgg 844  
DB 74 ATGAGGTNATTCAGGTGGG 56  
RESULT 7  
V61346/c  
ID V61346 standard; cDNA; 328 BP.  
XX AC V61346;

```
XX 06-JAN-1999 (first entry)
XX Forward DNA sequence of prostate tumour clone 8-b5.
XX Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
XX Homo sapiens.
XX WO9837093-A2.
XX 27-AUG-1998.
XX 25-FEB-1998; 98WO-US03492.
XX 09-FEB-1998; 98US-0020956.
XX 25-FEB-1997; 97US-0806099.
XX 01-AUG-1997; 97US-0904804.
XX (CORI-) CORIXA CORP.
XX Dillon DC, Xu J;
XX WPI; 1998-609886/51.
XX Polypeptides comprising immunogenic portions of prostate proteins -
XX used in a vaccine for the treatment of prostate cancer
XX Claim 12; Page 121; 130pp; English.
XX The present sequence is a DNA which encodes an immunogenic portion
XX of a prostate tumour protein. The encoded immunogen, or the DNA itself,
XX can be used as a vaccine for the treatment of prostate cancer. The DNA
XX was identified by analysis of a subtracted cDNA library obtained by
XX subtracting a prostate tumour cDNA expression library with a normal
XX tissue cDNA library.
XX Sequence 328 BP; 84 A; 73 C; 61 G; 106 T; 4 other;

Query Match 3.3%; Score 34.6; DB 19; Length 328;
Best Local Similarity 56.5%; Pred. No. 0.23;
Matches 61; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 627 ctaccacaagtggtgtaagctgataaactgctatatttattaaactggc 686
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 283 CTACCACCACCTGGCCCTNTCCAGAGAGTAAGAGTNGTAGGGGTAAATTTTCAATCTCTC 224
QY 687 tgactggtgctgtgtaagctgataaactgctatatttattaaactggc 734
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 223 TGGCTGATGATAAGGATCCCAAGTCAATGGGATGAAATTTNAACTGGC 176

RESULT 8
V58682/c
ID V58682 standard; cDNA; 328 BP.
XX V58682;
XX 08-DEC-1998 (first entry)
XX Prostate tumour specific gene clone 8-b5.
XX Prostate tumour specific gene; human; prostate cancer; detection;
XX therapy; ss.
XX Homo sapiens.
XX WO9837418-A2.
XX 27-AUG-1998.
XX 25-FEB-1998; 98WO-US03690.
XX (CORI-) CORIXA CORP.
```

```
XX 09-FEB-1998; 98US-0904809.
XX 25-FEB-1997; 97US-0806596.
XX 01-AUG-1997; 97US-0904809.
XX (CORI-) CORIXA CORP.
XX Dillon DC, Xu J;
XX WPI; 1998-480803/41.
XX Novel human prostate specific tumour protein and fragments - useful
XX for detecting and treating prostate cancers
XX Claim 1; Page 130; 141pp; English.
XX This sequence represents a human prostate tumour specific gene, and can
XX be used in the method of the invention. The method is for detecting
XX prostate cancer; comprises contacting a biological sample with an agent
XX able to bind an immunogenic portion of a prostate protein (such as
XX encoded by this sequence). An antibody which binds to an immunogenic
XX portion of the prostate protein, and the method can be used to detect,
XX monitor progression of, or treat prostate cancers. The antibody may
XX also be conjugated to a therapeutic agent for use in therapy of prostate
XX cancers.
XX Sequence 328 BP; 84 A; 73 C; 61 G; 106 T; 4 other;

Query Match 3.3%; Score 34.6; DB 19; Length 328;
Best Local Similarity 56.5%; Pred. No. 0.23;
Matches 61; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 627 ctaccacaagtggtgtaagctgataaactgctatatttattaaactggc 686
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 283 CTACCACCACCTGGCCCTNTCCAGAGAGTAAGAGTNGTAGGGGTAAATTTTCAATCTCTC 224
QY 687 tgactggtgctgtgtaagctgataaactgctatatttattaaactggc 734
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 223 TGGCTGATGATAAGGATCCCAAGTCAATGGGATGAAATTTNAACTGGC 176

RESULT 9
A06445/c
ID A06445 standard; cDNA; 328 BP.
XX A06445;
XX 13-JUN-2000 (first entry)
XX Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:212.
XX Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
XX immunogenic; cyostatic; vaccine; ss.
XX Homo sapiens.
XX WO200004149-A2.
XX 27-JAN-2000.
XX 14-JUL-1999; 99WO-US15838.
XX 14-JUL-1998; 98US-0115453.
XX 14-JUL-1998; 98US-0116134.
XX 23-SEP-1998; 98US-0159812.
XX 23-SEP-1998; 98US-0159822.
XX 15-JAN-1999; 99US-0232149.
XX 15-JAN-1999; 99US-0232880.
XX 09-APR-1999; 99US-0288946.
XX (CORI-) CORIXA CORP.
```







SQ Sequence 762 BP; 230 A; 171 C; 164 G; 190 T; 7 other;

Query Match 3.2%; Score 33.4; DB 21; Length 762;  
Best Local Similarity 54.5%; Pred. No. 0.91;  
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

779	gtgacatattttggaagatgaagctctttatagagatgcttaagttttaaacgagacatttaa	838
Qy		
468	gtgactatatttttggaaacaggggcctttaaagagacagttaagtggaaaaggagcccttag	527
Db		
839	agcggcgtctattccatttaatgaatgggtgcctcaaaaggagaacactggagacaggg	898
Qy		
528	tatggcctcagtttaattctgacacagcccttatcagatttaataaagtttaatacacagaaa	587
Db		

RESULT 15  
A34785/c  
ID A34785 standard; DNA: 2030 BP.

AC A34785;

DT 28-JUL-2000 (first entry)

Human adenosine receptor related polynucleotide SEQ ID NO:2474.

Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

XX  
PN  
WO200009525-A2.

PD 24-FEB-2000.

AA  
PF 03-AUG-1999; 99WO-US17712.

03-AUG-1998: 98US-0095212.

PA (UYEC-) UNIV EAST CAROLINA.

XX  
PT  
Nvce JW:

XX  
DR WPI: 2000-205971/18.

xx New antisense oligonucleotides useful for treating e.g. pulmonary  
PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
PT cancers -

PS Disclosure: Page 609: 1343pp: English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 1%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:09:28 ; Search time 141.08 seconds  
(without alignments)  
1134.204 Million cell updates/sec

Title: US-09-101-423A-2  
Perfect score: 1058  
Sequence: 1 attgtgtgagcctattagc.....acacattggaacgttggtc 1058

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	39	3.7	437	6	PCT-US95-16766-29
C 2	39	3.7	7218	1	US-08-232-463-14
C 3	31.8	3.0	2026	4	US-08-993-228-3
C 4	31.2	2.9	2513	1	US-07-686-322A-2
C 5	31.2	2.9	2513	1	US-08-008-999-2
C 6	30.2	2.9	1700	6	PCT-US92-02091-1
C 7	30.2	2.9	1814	2	US-08-579-667-7
C 8	29.8	2.8	1278	4	US-08-853-659A-13
C 9	29.8	2.8	8119	2	US-08-460-343B-1
C 10	29.8	2.8	8119	2	US-08-398-028B-1
C 11	29.8	2.8	8119	3	US-08-504-285B-1
C 12	29.8	2.8	15512	4	US-08-853-659A-5
C 13	29.8	2.8	15512	4	US-08-853-659A-8
C 14	29.8	2.8	15512	4	US-08-853-659A-63
C 15	29.8	2.8	15512	4	US-08-853-659A-66
C 16	29.8	2.8	24701	4	US-08-853-659A-2
C 17	29.8	2.8	24701	4	US-08-853-659A-3
C 18	29.8	2.8	24701	4	US-08-853-659A-60
C 19	29.8	2.8	24701	4	US-08-853-659A-61
C 20	29.6	2.8	2886	4	US-08-687-080-55
C 21	29.6	2.8	3158	3	US-08-464-517-36
C 22	29.6	2.8	3158	4	US-08-246-361A-36
C 23	29.6	2.8	3158	5	US-08-463-772-36
C 24	29.6	2.8	3846	4	US-08-845-161A-5
C 25	29.4	2.8	3704	4	US-09-014-969-20
C 26	29.4	2.8	3905	4	US-08-666-271-1

27 29.2 2.8 2088 3 US-08-602-264A-1 Sequence 1, Appl  
28 29.2 2.8 2088 5 US-08-461-018A-1 Sequence 1, Appl  
29 29.2 2.8 10207 2 US-08-920-812-2 Sequence 2, Appl  
30 29.2 2.8 10207 2 US-08-920-827-2 Sequence 2, Appl  
31 29.2 2.8 10207 2 US-08-921-177-2 Sequence 2, Appl  
32 29.2 2.8 10207 2 US-08-362-577C-2 Sequence 2, Appl  
33 29.2 2.8 10207 3 US-08-920-828-2 Sequence 2, Appl  
34 29.2 2.8 87350 5 US-08-781-891-79 Sequence 79, Appl  
35 29.2 2.7 3635 3 US-08-553-436A-5 Sequence 5, Appl  
36 28.8 2.7 1827 4 US-08-737-371A-3 Sequence 3, Appl  
37 28.8 2.7 1827 6 PCT-US95-05853-3 Sequence 3, Appl  
38 28.8 2.7 21126 1 US-08-008-216-19 Sequence 19, Appl  
39 28.8 2.7 21126 1 US-08-459-569-19 Sequence 19, Appl  
40 28.8 2.7 21126 2 US-08-458-831-19 Sequence 19, Appl  
41 28.6 2.7 2223 2 US-08-257-073-4 Sequence 4, Appl  
42 28.6 2.7 2433 1 US-08-248-021A-1 Sequence 1, Appl  
43 28.6 2.7 2799 1 US-08-212-188-1 Sequence 1, Appl  
44 28.6 2.7 2799 5 US-08-970-725-1 Sequence 1, Appl  
45 28.6 2.7 2799 6 PCT-US95-02708-1 Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
PCT-US95-16766-29/c  
Sequence 29, Application PC/TUS9516766  
GENERAL INFORMATION:  
APPLICANT: COLD SPRING HARBOR, Laboratories  
TITLE OF INVENTION: CANCER DETECTION PROBES  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16766  
FILING DATE: 12-DEC-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: ROWLAND, Bertram I  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: FP60430-1-PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "probe"  
PCT-US95-16766-29

Query Match 3.7%; Score 39; DB 6; Length 437;  
Best Local Similarity 59.5%; Pred. No. 0.0059;  
Matches 66; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 760 aaaaaaaaaaacacatttaggtgacatttttgagatgaagatgcttaagatgcttaa 819  
Db 307 AATATATATACCTGTAATGTCCTTTATTGGAAATAGGTCCTTTCACAGATGATTCAA 248







COUNTRY: U.S.A.  
ZIP: 99352  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 3.50", 1.44 Mb storage  
COMPUTER: IBM PC/XT/AT  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Processor (WordPerfect 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/853,659A  
FILING DATE: Unknown  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: none  
FILING DATE: n/a  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1278 bases  
TYPE: nucleotide  
STRANDEDNESS: double stranded  
TOPOLOGY: linear  
FEATURE:

OTHER INFORMATION: the coding nucleotides of SEQ ID  
NO:13 correspond to nucleotides 4350 through  
OTHER INFORMATION: 5627 of SEQ ID NO:2  
US-08-853-659A-13

Query Match 2.8%; Score 29.8; DB 4; Length 1278;  
Best Local Similarity 50.3%; Pred. No. 8.6;  
Matches 73; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 722 ttataaactggcttttttgaacacccactcaaggaaacacacacacttagggtg 781  
DB 711 TAAGATAGTCGCAATTCATTATTAAGCGCTCAATATCATTAACACCACTCAATATC 652  
QY 782 acattatttgagatgaagcttttatagagatgcttaagtttaaacgagacatttaaagc 841  
DB 651 ATCTTTTTTAAGTGATGCTACITTCATATCATTAATTTTCAGACITTAACITTTATATA 592  
QY 842 cggctctattccatttaataatgaatg 866  
DB 591 CGCCTGTTCTTGTGAAGAAAATTG 567

RESULT 9  
US-08-460-343B-1  
Sequence 1, Application US/08460343B  
Patent No. 5741664  
GENERAL INFORMATION:  
APPLICANT: Marcus D. Ballinger and James A. Wells  
TITLE OF INVENTION: SUBSTITUTED VARIANTS CAPABLE OF CLEAVING  
TITLE OF INVENTION: SUBSTRATES CONTAINING DIBASIC RESIDUES  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,343B  
FILING DATE: 01-Jun-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/398028  
FILING DATE: 03-mar-1995  
ATTORNEY/AGENT INFORMATION:

NAME: Kubinec, Jeffrey S.  
REGISTRATION NUMBER: 36,575  
REFERENCE/DOCKET NUMBER: P0936C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-8228  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8115 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-460-343B-1

Query Match 2.8%; Score 29.8; DB 2; Length 8119;  
Best Local Similarity 47.2%; Pred. No. 24;  
Matches 91; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 671 aagacatttgcgtgactggtgctgtaagctgataaactgctctattattataaac 730  
DB 6905 AAGGATTTTCCTAGCTCAATCTTTTAAAAAACACAAAGACCATTTTTTAATG 6964  
QY 731 tggcttttttgaacacccactcaaggaaacacacacacttagggtgacattatt 790  
DB 6965 TGGTCTTTTATCTTCAACTTAAAGCACCCTATTAGTTTCAACAAAGAAATTTGATAAGTG 7024  
QY 791 ggagatgaagttttatagagatgcttaagtttaaacgagacgttttaaacgcggctctat 850  
DB 7025 GGATATTTTAAAAATATATATTTATGTTACAGTAATATTTGACTTTTAAAAAAGGATTGAT 7084  
QY 851 tccatttaaatgaa 863  
DB 7085 TCTAATCAAGAAA 7097

RESULT 10  
US-08-398-028B-1  
Sequence 1, Application US/08398028B  
Patent No. 5780285  
GENERAL INFORMATION:  
APPLICANT: Marcus D. Ballinger and James A. Wells  
TITLE OF INVENTION: SUBSTITUTED VARIANTS CAPABLE OF CLEAVING  
TITLE OF INVENTION: SUBSTRATES CONTAINING DIBASIC RESIDUES  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/398,028B  
FILING DATE: 03-Mar-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kubinec, Jeffrey S.  
REGISTRATION NUMBER: 36,575  
REFERENCE/DOCKET NUMBER: P0936  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-8228  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8119 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2000, 22:07:41 ; Search time 3273.53 Seconds  
(without alignments)  
1998.274 Million cell updates/sec

Title: US-09-101-423A-2  
Perfect score: 1058  
Sequence: 1 attgcgtgagcctattagc.....acacattogaacgcttggtc 1058

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estom:\*  
68: em\_estov1:\*  
69: em\_estov2:\*  
70: em\_estpl1:\*  
71: em\_estpl2:\*  
72: em\_estpl3:\*  
73: em\_estpl4:\*  
74: em\_estpl5:\*  
75: em\_estro1:\*  
76: em\_estro2:\*  
77: em\_estro3:\*  
78: em\_estro4:\*  
79: em\_estro5:\*  
80: em\_estro6:\*  
81: em\_estro7:\*  
82: em\_estro8:\*  
83: em\_estro9:\*  
84: em\_estro10:\*  
85: em\_estro11:\*  
86: em\_estro12:\*  
87: em\_estro13:\*  
88: gb\_gss1:\*  
89: gb\_gss2:\*  
90: gb\_gss3:\*  
91: gb\_gss4:\*  
92: em\_gss1:\*  
93: em\_gss2:\*  
94: em\_gss3:\*  
95: em\_gss4:\*  
96: gb\_gss5:\*  
97: gb\_gss6:\*  
98: gb\_gss7:\*  
99: gb\_gss8:\*  
100: gb\_gss9:\*  
101: em\_gss5:\*  
102: em\_gss6:\*  
103: em\_gss7:\*  
104: em\_gss8:\*  
105: em\_gss9:\*  
106: em\_gss10:\*  
107: em\_gss11:\*  
108: gb\_gss10:\*  
109: gb\_gss11:\*  
110: em\_gss12:\*  
111: gb\_gss12:\*  
112: gb\_gss13:\*  
113: gb\_gss14:\*  
114: gb\_gss15:\*  
115: gb\_gss16:\*  
116: gb\_gss17:\*



```

QY 643 ttccaaagacacacatatattggagcaagacattttgtgctgactgactgtgta 702
DB 137 TTTCCAAAGACACACATACATTTGGGCAAGACATTTTGTGCTGCTGCTGTGTA 78
QY 703 agctgataaactgtattattataaactgcttttttttgaacacccccactcaagaaa 762
DB 77 AGCTGATAAATGCTATATTTATTAATACTGCTTTTCTTGAACACTTCATTGAGGAAA 18
QY 763 aaaaaacacacttaggg 779
DB 17 ACCCCCCACAGTGAGG 1

RESULT 2
LOCUS AI425066 311 bp mRNA EST 30-MAR-1999
DEFINITION tg50h03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2112245 3', mRNA sequence.
ACCESSION AI425066
VERSION AI425066.1 GI:4270997
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 382 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES
source
1..311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2112245"
/clone.lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7R3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 97 a 62 c 55 g 97 t
ORIGIN
Query Match 14.3%; Score 151.2; DB 10; Length 311;
Best Local Similarity 76.7%; Pred. No. 8.1e-34;
Matches 20; Conservative 0; Mismatches 53; Indels 8; Gaps 1;

QY 218 ccaggagactctgtagaaataacatcagaccatgaggaggaattgatcatcatt----- 271
DB 5 CTACAGATTATTATAGAAACAGCATCAGACAATGAAACCTGACAAATATTATTCGGA 64
QY 272 --gttttcaatgggtatgcgcaagggaattccatctgattataaaataactactgctg 329
DB 65 TGTTACATGCTGCTGATCATCAAGGGAACCTTGCAATGATTAATAAATAATTGCTG 124
QY 330 gcactaaatccaatggaaatgccccacacaaatttatcttccattcatgctgctaccat 389

```

```

DB 125 GCATTAATCCAGTTGGATCACCACACGATTTATCTTCTATTTCATGCTGTACCAT 184
QY 390 atgcctgacgtggcgagcagaagcattccctccgttctctgataaataagctactttgtaaa 449
DB 185 ATGCCTGACATTGTGTACAGAAACATTTCTTTCATTCTGTATAAGTAGTACTTTGTAAA 244
QY 450 tatttggagacgggagctctgg 471
DB 245 TATTAGAGCGGAGGACTCTGG 266

RESULT 3
LOCUS B48647 449 bp DNA GSS 08-APR-1999
DEFINITION RPC111-2M21-TV RPC111 Homo sapiens genomic clone RPC111-2M21, DNA
sequence.
ACCESSION B48647
VERSION B48647
KEYWORDS B48647.1 GI:2600884
SOURCE GSS.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 449)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter
,J.C.
TITLE Use of EAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other-GSSs: RPC111-2M21.TP
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC111. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (inforesgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html
Seq primer: T7
Class: EAC ends.
FEATURES
source
1..449
/organism="Homo sapiens"
/db_xref="GDB:7500692"
/db_xref="taxon:9606"
/clone="RPC111-2M21"
/clone.lib="RPC111"
/sex="Male"
/cell_type="Lymphocytes"
/vector="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT 144 a 102 c 112 g 91 t
ORIGIN
Query Match 4.9%; Score 51.6; DB 118; Length 449;
Best Local Similarity 68.0%; Pred. No. 0.00023;
Matches 87; Conservative 0; Mismatches 39; Indels 2; Gaps 1;

QY 780 tgacattatttggagatgaagtcttttatagagatgcttaagtttaacagagacttttaa 839
DB 47 TGACCTTATTGGAGTAGGTCTCTTACAGAGGTAAATCAAGTTAAATGAGGTCATTAGG 106
QY 840 gccggctctattccatttaagtgtccctacaaaggagaaactgggacaggggt 899
DB 107 GGGGACCCCTATGCAATC--TGACTGTGTGTCCTTACAAAGGGAGAAATCTGGACACAG 164

```

```

QY 900 atgtacac 907
| | | | |
Db 165 ACATGCAC 172

RESULT 4
B51686/c
LOCUS C1978SK-A-991G9.TP C1978SK Homo sapiens genomic clone A-991G9, 18-JUN-1998
DEFINITION DNA sequence.
ACCESSION B51686
VERSION B51686.1 GI:2603923
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 479)
Kim, U.-J., Adams, M.D. and Simon, M.I.
Determination of clone end sequences of human Bacterial Artificial
Chromosomes
Unpublished (1997)
Other_GSSs: C1978SK-A-991G9.TV
Contact: Ung-Jin Kim
Caltech Genome Research Lab
California Institute of Technology
Division of Biology, MS 147-75, Pasadena, CA 91125, USA
Tel: 626 796 7066
Fax: 626 395 4901
Email: ung@ash.tree.caltech.edu
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..479
/organism="Homo sapiens"
/db_xref="GDB:5203290"
/db_xref="taxon:9606"
/clone="A-991G9"
/clone_lib="C1978SK"
/sex="Female"
/cell_type="Fibroblast"
/notes="Vector: pBAC1081; Site_1: HindIII; Site_2: HindIII;
Caltech Human BAC Library A"
BASE COUNT 127 a 103 c 84 g 165 t
ORIGIN

FEATURES
source
Location/Qualifiers
1..479
/organism="Homo sapiens"
/db_xref="GDB:5203290"
/db_xref="taxon:9606"
/clone="A-991G9"
/clone_lib="C1978SK"
/sex="Female"
/cell_type="Fibroblast"
/notes="Vector: pBAC1081; Site_1: HindIII; Site_2: HindIII;
Caltech Human BAC Library A"

Query Match 4.8%; Score 50.4; DB 118; Length 479;
Best Local Similarity 71.1%; Pred. No. 0.00053;
Matches 81; Conservative 0; Mismatches 31; Indels 2; Gaps 1;

QY 776 aggtgacattattgagatgaagctttatagagatgcttaagtttaaacgagacttt 835
| | | | |
Db 205 AGTGTGACCTATTGAGATGAAGGTCCTACAGAGTAATCAAGTTAAATGAGGTGT 146
| | | | |

QY 836 taaagccggctctattccattgaatggtgtccctacaaaggagaactg 889
| | | | |
Db 145 TACAATGTCCTAATCCAGT--ATGACTGTTGCTCCATATAAAGAGGAATG 94
| | | | |

RESULT 5
B70626/c
LOCUS CIT-HSP-2059P21.TF CIT-HSP Homo sapiens genomic clone 2059P21, DNA
DEFINITION sequence.
ACCESSION B70626
VERSION B70626.1 GI:2709850
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 561)
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other_GSSs: RPC111-46G20.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200

Query Match 4.8%; Score 50.4; DB 118; Length 479;
Best Local Similarity 71.1%; Pred. No. 0.00053;
Matches 81; Conservative 0; Mismatches 31; Indels 2; Gaps 1;

QY 776 aggtgacattattgagatgaagctttatagagatgcttaagtttaaacgagacttt 835
| | | | |
Db 205 AGTGTGACCTATTGAGATGAAGGTCCTACAGAGTAATCAAGTTAAATGAGGTGT 209
| | | | |

QY 836 taaagccggctctattccattgaatggtgtccctacaaaggagaactg 889
| | | | |
Db 208 TACAATGTCCTAATCCAGT--ATGACTGTTGCTCCATATAAAGAGGAATG 157
| | | | |

RESULT 6
AQ201082
LOCUS RPC111-46G20.TK RPC111 Homo sapiens genomic clone RPC111-46G20,
DEFINITION DNA sequence.
ACCESSION AQ201082
VERSION AQ201082.1 GI:3613281
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 281)
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other_GSSs: RPC111-46G20.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200

Query Match 4.8%; Score 50.4; DB 118; Length 561;
Best Local Similarity 71.1%; Pred. No. 0.00056;
Matches 81; Conservative 0; Mismatches 31; Indels 2; Gaps 1;

QY 776 aggtgacattattgagatgaagctttatagagatgcttaagtttaaacgagacttt 835
| | | | |
Db 268 AGTGTGACCTATTGAGATGAAGGTCCTACAGAGTAATCAAGTTAAATGAGGTGT 209
| | | | |

QY 836 taaagccggctctattccattgaatggtgtccctacaaaggagaactg 889
| | | | |
Db 208 TACAATGTCCTAATCCAGT--ATGACTGTTGCTCCATATAAAGAGGAATG 157
| | | | |

BASE COUNT 145 a 122 c 111 g 183 t
ORIGIN

FEATURES
source
Location/Qualifiers
1..561
/organism="Homo sapiens"
/db_xref="GDB:7060224"
/db_xref="taxon:9606"
/clone="2059P21"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/notes="Vector: pBAC1081; Site_1: HindIII; Site_2:
HindIII"

```







DB 432 GTGATGATTATGAGATAAGAGCTTTAAAGAGGTAATTAAGTTAATAGGTCATTAG 373  
 QY 839 agccggctctattccatttaataatgaatggtgtccctacaaag 879  
 DB 372 GGTAGGCGCTGAATCCATCAAGACTGGTATCCTCTAATAAG 332

RESULT 11  
 A0629762 298 bp DNA GSS 17-JUN-1999  
 LOCUS  
 DEFINITION RPCI-11-467D3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-467D3,  
 DNA sequence.  
 ACCESSION A0629762  
 VERSION A0629762.1 GI:5092397  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 298)  
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter  
 ,J.C.  
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
 Map Building  
 JOURNAL Unpublished (1997)  
 COMMENT Other\_GSSs: RPCI-11-467D3.TJ  
 Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbeetigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
 Research Genet cs (info@resgen.com). BAC end search page:  
 http://www.tigr.org/tldb/humgen/bac\_end\_search.html.  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..298  
 /organism="Homo sapiens"  
 /db\_xref="GB:7679018"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-467D3"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPCI11 Human Male BAC Library"  
 BASE COUNT 93 a 46 c 81 g 78 t  
 ORIGIN

Query Match 4.5%; Score 47.2; DB 100; Length 298;  
 Best Local Similarity 59.8%; Pred. No. 0.0041;  
 Matches 79; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 776 aggtgacattatttgagatgaagtctttatagagatgcttaagttaaagagacttt 835  
 DB 167 AGTGGGGCCCTATTGGAGATAAGGCTTTATAGAGGCAATCAAGTTAAATGAGTCAAT 226  
 QY 836 taaagcgggtctattccatttaataatgaatggtgtccctacaaaggaagaaactgggacag 895  
 DB 227 TAGGCGAGGCCCTAATCATATATAGTACTGTCTCTATATAAAGGGGAAACITGGCCAGG 286

QY 896 aggtatgtacac 907  
 DB 287 TGGATGGCTCAC 298

RESULT 12  
 A0339795 445 bp DNA GSS 19-MAY-1999  
 LOCUS  
 DEFINITION RPCI-11-316E14.TV RPCI-11 Homo sapiens genomic clone RPCI-11-316E14  
 , DNA sequence.  
 ACCESSION A0339795  
 VERSION A0339795.1 GI:4870325  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 445)  
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter  
 ,J.C.  
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
 Map Building  
 JOURNAL Unpublished (1997)  
 COMMENT Other\_GSSs: RPCI-11-316E14.TJ  
 Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbeetigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
 Research Genet cs (info@resgen.com). BAC end search page:  
 http://www.tigr.org/tldb/humgen/bac\_end\_search.html.  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 source  
 Location/Qualifiers  
 1..445  
 /organism="Homo sapiens"  
 /db\_xref="GB:7621069"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-316E14"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPCI11 Human Male BAC Library"  
 BASE COUNT 136 a 85 c 101 g 123 t  
 ORIGIN

Query Match 4.4%; Score 46.8; DB 98; Length 445;  
 Best Local Similarity 69.6%; Pred. No. 0.0061;  
 Matches 78; Conservative 0; Mismatches 32; Indels 2; Gaps 1;

QY 779 gtgacattatttgagatgaagtctttatagagatgcttaagttaaagagactttta 838  
 DB 81 GTGACTGATTATGGACATAGGCGCTTTAAAGAGCTGATTAGATTAATGAGGCTGTGAG 140  
 QY 839 agccggctctattccatttaataatgaatggtgtccctacaaaggaagaaactgg 890  
 DB 141 GGTGGGCCCTAATCCAATC--TGACTGTGTCTTCTTATAAGAGAACATATTGG 190

RESULT 13  
 A0201078 276 bp DNA GSS 20-APR-1999  
 LOCUS  
 DEFINITION RPCI11-46G18.TK RPCI-11 Homo sapiens genomic clone RPCI-11-46G18,  
 DNA sequence.  
 ACCESSION A0201078  
 VERSION A0201078.1 GI:3613277  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:46:23 ; Search time 7111.01 Seconds  
(without alignments)  
619.113 Million cell updates/sec

Title: us-09-101-423a-3

Perfect score: 1008  
Sequence: 1 aggaccagagttcacatccc.....ttccctgttccctcctggaac 1008

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1033670 segs, 2183789903 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_bal:\*  
2: gb\_bal:\*  
3: gb\_bal:\*  
4: gb\_bal:\*  
5: gb\_bal:\*  
6: gb\_bal:\*  
7: gb\_bal:\*  
8: gb\_bal:\*  
9: gb\_bal:\*  
10: gb\_bal:\*  
11: gb\_bal:\*  
12: gb\_bal:\*  
13: gb\_bal:\*  
14: gb\_bal:\*  
15: gb\_bal:\*  
16: gb\_bal:\*  
17: gb\_bal:\*  
18: gb\_bal:\*  
19: gb\_bal:\*  
20: gb\_bal:\*  
21: gb\_bal:\*  
22: gb\_bal:\*  
23: gb\_bal:\*  
24: gb\_bal:\*  
25: gb\_bal:\*  
26: gb\_bal:\*  
27: gb\_bal:\*  
28: gb\_bal:\*  
29: gb\_bal:\*  
30: gb\_bal:\*  
31: gb\_bal:\*  
32: gb\_bal:\*  
33: gb\_bal:\*  
34: gb\_bal:\*  
35: gb\_bal:\*  
36: gb\_bal:\*  
37: gb\_bal:\*  
38: gb\_bal:\*  
39: gb\_bal:\*  
40: gb\_bal:\*  
41: gb\_bal:\*  
42: gb\_bal:\*  
43: gb\_bal:\*

44: em\_hum5:\*  
45: gb\_bal:\*  
46: gb\_bal:\*  
47: gb\_bal:\*  
48: gb\_bal:\*  
49: gb\_bal:\*  
50: gb\_bal:\*  
51: gb\_bal:\*  
52: gb\_bal:\*  
53: gb\_bal:\*  
54: gb\_bal:\*  
55: gb\_bal:\*  
56: gb\_bal:\*  
57: gb\_bal:\*  
58: gb\_bal:\*  
59: gb\_bal:\*  
60: gb\_bal:\*  
61: gb\_bal:\*  
62: gb\_bal:\*  
63: gb\_bal:\*  
64: gb\_bal:\*  
65: gb\_bal:\*  
66: gb\_bal:\*  
67: gb\_bal:\*  
68: gb\_bal:\*  
69: gb\_bal:\*  
70: gb\_bal:\*  
71: gb\_bal:\*  
72: gb\_bal:\*  
73: gb\_bal:\*  
74: gb\_bal:\*  
75: gb\_bal:\*  
76: gb\_bal:\*  
77: gb\_bal:\*  
78: gb\_bal:\*  
79: gb\_bal:\*  
80: gb\_bal:\*  
81: gb\_bal:\*  
82: gb\_bal:\*  
83: gb\_bal:\*  
84: gb\_bal:\*  
85: gb\_bal:\*  
86: gb\_bal:\*  
87: gb\_bal:\*  
88: gb\_bal:\*  
89: gb\_bal:\*  
90: gb\_bal:\*  
91: gb\_bal:\*  
92: gb\_bal:\*  
93: gb\_bal:\*  
94: gb\_bal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1008	1008	5 A64060	A64060 Sequence 3
2	53	5.3 17488	63 AC078885	AC078885 Mus muscu
3	50	5.0 208655	63 AC077689	AC077689 Mus muscu
4	50	5.0 215795	55 AC067964	AC067964 Mus muscu
5	49	4.9 39751	50 AC025910	AC025910 Mus muscu
6	49	4.9 45966	14 AP000028	AP000028 Homo sapi
7	49	4.9 201251	50 AC025964	AC025964 Mus muscu
8	49	4.9 207757	49 AC023934	AC023934 Homo sapi
9	49	4.9 267762	57 AC073817	AC073817 Mus muscu
10	48	4.8 212481	51 AC026760	AC026760 Mus muscu
11	46	4.6 118	88 AF168591	AF168591 Rattus norv
12	46	4.6 1201	88 R012268	R012268 Rattus norv

AF095690	Mus muscu	QY	121	agaacacatcacaaataaaaaaactctgaaaaatttaagctaaattgtaagaat	180
AF01577	Mouse gene				
AF246667	Mus muscu	Db	121	AGAACACAAATCAAAATAAAAAATCTGAAATATTTAAAGCTAAATGTTAAGAAAT	180
U57691	Mus muscu	QY	181	aacatatacaattttcttatattttttaaagattatttttatatatagta	240
AC003995	Mouse Cos				
Continuation (4 of					
AC007665	Mus muscu	Db	181	AACATATTAACAATTTTCTTTATTTTAAAAATTTATTTTATATATATAGTA	240
AC065520	Mus muscu				
Continuation (2 of					
AC073667	Mus muscu	QY	241	caatgcctctccctccagacataagcagtlacaggcatalcgatcccatlaacagatgltgt	300
AC069016	Mus muscu	Db	241	CACAGCCTCTCCCTCCAGACATAGCAATACAGGGCATCGGATCCCAATTACAGATGGTTGT	300
AC073709	Mus muscu	QY	301	gagccacacatggtgttccaagaatggtgtgagagccacacatggtgttcaagaaattgaact	360
AC025053	Mus muscu				
AF027865	Mus muscu	Db	301	GAGCCACCAATGTGGTTTCACAGATGGTTGTGAGCCACACATGTGGTTTCAAGAAATTGAAC	360
AC066447	Mus muscu				
AC006404	Mus muscu	QY	361	caggacaccttgyaagagcagtcagtgctcttaacctcctaagcactctcctgaacctta	420
AC006404	Mus muscu	Db	361	CAGGACCTTTGSAAGAGCACTAGTGCCTTTAACCTCTAAGCCATGCTCTCAACCTTA	420
AC069560	Mus muscu				
AC021952	Mus muscu	QY	421	tatcaaatcttaactgctcagctacacacaaattctcttctcttaatggtgagatttttg	480
AC074359	Mus muscu	Db	421	TATCAATTTTAATGCTACCTACACACAACTTCTCTTTCTTTAAAGTTGAATTTTGG	480
AC074359	Mus muscu				
AC004093	Mus muscu	QY	481	tctgyagaagaaagaataaagagaggaagaaacatgctcttcacattgcacagstggaaa	540
AC009296	Mus muscu	Db	481	TCGTGAGAACGTAAAGATTAAGAGGAGGGAAGAAACATTTGCTTTCACATTGCACCACTGGGAA	540
AF111103	Mouse maj	QY	541	cagcgtgttttaaaagttaggaatgccatgaaatgacatgctggctcttcattactgttctc	600
AC073848	Mus muscu	Db	541	CAGCGTGTTTAAAGTAGGAATGCCAATGAATGACTGGCTGCTCTCTCATTTACTGTTCTT	600
AC008246	Homo sap1				
AC068007	Mus muscu	QY	601	ccacactccctcttaactcagagctccctctatcttaatttataatgattagcagatcccgaa	660
AF220294	Mus muscu				
AC068627	Mus muscu	Db	601	ccacactccctcttaactcagagctccctctatcttaatttataatgattagcagatcccgaa	660
AC046146	Mus muscu				
AC041194	Mus muscu	QY	601	ccacactccctcttaactcagagctccctctatcttaatttataatgattagcagatcccgaa	660
AC023301	Mus muscu				

29-MAR-1995

BASE COUNT ORIGIN	274 a	227 c	201 g	306 t
----------------------	-------	-------	-------	-------

```
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

61 AGGGGCGCACTGCACACACAGTGCACATCACTTACAGAGACACACACATTCAGCAGCATA 120

Db 961 CAGAGCTGAAATTCCTGTGCTCTTCCCTGTCTCCTCGAAC 1008

Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Rodentia; Sclurognathi;  
Mammalia; Eutheria; Muridae; Mus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
COMMENT

1 (bases 1 to 17488)  
Montgomery K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Joshikhes,I.P., Shum,C., Decker,J., Thomas,E., Pereira,A.,  
Gordon,M., Goltz,V.S. and Kucherlapati,R.  
High Throughput Mouse Sequencing  
Unpublished  
2 (bases 1 to 17488)  
Montgomery K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Joshikhes,I.P., Shum,C., Decker,J., Thomas,E., Pereira,A.,  
Gordon,M., Goltz,V.S. and Kucherlapati,R.  
Direct Submission  
Submitted (08-JUG-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA  
-----Genome Center Center:  
Albert Einstein College of Medicine Center  
Code: AECOM  
Web site:  
http://sequence.aecom.yu.edu/cg1-  
bin/ws\_exe/mousedB/mouseSeq/mouseseqtable.hns  
Contact: jhan@sequence.aecom.yu.edu  
-----Summary Statistics  
Center project name: AAF  
Sequencing vector: pUC18; L08752  
Chemistry: Dye-terminator Big Dye; 100\*\*Consensus quality: 9471 at least Q20  
\*Consensus quality: 8035 at least Q30  
\*Consensus quality: 6349 at least Q40  
Estimated insert size: agarose-PP - N/A  
\*\*Estimated insert size: 17168 - sum-of-ctrls  
Quality coverage: agarose-PP - N/A  
Quality coverage: 2x sum-of-ctrls - N/A  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
971: contig of 971 bp in length  
972  
991: gap of unknown length  
992  
1923: contig of 932 bp in length  
1924  
1943: gap of unknown length  
1944  
2914: contig of 971 bp in length  
2915  
2934: gap of unknown length  
2935  
3920: contig of 986 bp in length  
3921  
3940: gap of unknown length  
3941  
4881: contig of 941 bp in length  
4901  
4901: gap of unknown length  
4902  
5892: contig of 991 bp in length  
5912  
5912: gap of unknown length  
5913  
6917: contig of 1005 bp in length  
6937  
6937: gap of unknown length  
6938  
7915: contig of 978 bp in length  
7916  
7935: gap of unknown length  
7936  
8929: contig of 994 bp in length  
8930  
8949: gap of unknown length  
8950  
9901: contig of 952 bp in length  
9902  
9921: gap of unknown length  
9922  
10895: contig of 974 bp in length  
10896  
10915: gap of unknown length  
10916  
11837: contig of 922 bp in length  
11838  
11857: gap of unknown length  
11858  
12843: contig of 986 bp in length  
12844  
12863: gap of unknown length  
12864  
13831: contig of 968 bp in length  
13832  
13851: gap of unknown length  
13852  
14970: contig of 1119 bp in length  
14971  
14990: gap of unknown length  
14991  
15890: contig of 900 bp in length  
15891  
15910: gap of unknown length

```

FEATURES
Source
155:11 17488: contig of 1578 bp in length.
Location/Qualifiers
1. 17488
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="6"
/clone="Rp23-117123"
/sex="male"

BASE COUNT 4497 a 3817 c 3856 g 4289 t 1029 others
ORIGIN

Query Match 5.3%; Score 53; DB 63; Length 17488;
Best Local Similarity 100.0%; Pred. No. 1.1e-15;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Oy 354 ttgaactcaagacaccttggagagacagtcagtccttcaactcaagaccatc 406
|||||
Db 1263 TTGAACCTCAGACACCTTTGGAAGACGACGTCAGTCCTTAACTCTAAGGCATC 1211

RESULT 3
AC077689 DNA HTG 02-AUG-2000
LOCUS Mus musculus clone Rr23-301H20 strain C57BL6/J. ** SEQUENCING IN
DEFINITION PROGRAMS ***, 41 unordered pieces.
AC077689
AC077689.1 GI:9653117
HTG: HTCS_PHASE1.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
1 (bases 1 to 208655)
Han,J., Montgomery,K.T., Gills,G., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Pereira,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 208655)
Han,J., Montgomery,K.T., Gills,G., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Pereira,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
Direct Submission
Submitted (02-AUG-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
-----Genome Center Center:
Albert Einstein College of Medicine Center
Code: AECCOM
Web site:
http://sequence.aecom.yu.edu/cg1-
bin/ws.exe/mousedb/mousedb/mouseseqtable.hts
Contact: jhan@sequence.aecom.yu.edu
-----Summary Statistics
Center project name: ACR
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 191517 at least Q20
*Consensus quality: 187352 at least Q30
*Consensus quality: 180858 at least Q40
*Estimated insert size: agarose-pp - N/A
**Estimated insert size: 207855 - sum-of-contigs
Quality coverage: agarose-pp - N/A
Quality coverage: 8.1x sum-of-contigs - N/A
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

* as soon as it is available and the accession number will	
* be preserved.	
1	49179: contig of 49179 bp in length
49180	49199: gap of unknown length
49200	67711: contig of 18012 bp in length
67212	67731: gap of unknown length
67232	77669: contig of 10638 bp in length
77870	77889: gap of unknown length
77890	85533: contig of 7944 bp in length
85834	85853: gap of unknown length
85854	94023: contig of 8170 bp in length
94024	94043: gap of unknown length
94044	103931: contig of 9858 bp in length
103902	103921: gap of unknown length
103922	110389: contig of 6468 bp in length
110390	110409: gap of unknown length
110410	118926: contig of 7817 bp in length
118227	118246: gap of unknown length
118247	126519: contig of 8273 bp in length
126520	126539: gap of unknown length
126540	134489: contig of 7950 bp in length
134490	134509: gap of unknown length
134510	139301: contig of 4192 bp in length
139302	139321: gap of unknown length
139322	143030: contig of 3709 bp in length
143031	143050: gap of unknown length
143051	146004: contig of 2954 bp in length
146005	146024: gap of unknown length
146025	152866: contig of 6942 bp in length
152967	152886: gap of unknown length
152987	156599: contig of 3973 bp in length
156960	156979: gap of unknown length
156980	159924: contig of 2945 bp in length
159925	159944: gap of unknown length
159945	164543: contig of 4599 bp in length
164544	164563: gap of unknown length
164564	166657: contig of 2094 bp in length
166658	166677: gap of unknown length
166678	169617: contig of 2940 bp in length
169618	169637: gap of unknown length
169638	171581: contig of 1944 bp in length
171582	171601: gap of unknown length
171602	175120: contig of 3519 bp in length
175121	175140: gap of unknown length
175141	177094: contig of 1934 bp in length
177095	177114: gap of unknown length
177115	181168: contig of 4354 bp in length
181169	181488: gap of unknown length
181489	184465: contig of 2977 bp in length
184466	184485: gap of unknown length
184886	187021: contig of 2336 bp in length
187022	187041: gap of unknown length
187042	188912: contig of 1871 bp in length
188913	188932: gap of unknown length
188933	192284: contig of 3352 bp in length
192285	192304: gap of unknown length
192305	193992: contig of 1688 bp in length
193993	194012: gap of unknown length
194013	194013: contig of 390 bp in length
194403	194422: gap of unknown length
194423	195331: contig of 909 bp in length
195332	195351: gap of unknown length
195352	196533: contig of 1182 bp in length
196534	196553: gap of unknown length
196554	197669: contig of 1316 bp in length
197670	197689: gap of unknown length
197870	199071: contig of 1182 bp in length
199072	199091: gap of unknown length
199092	200023: contig of 912 bp in length
200024	200023: gap of unknown length
200024	201636: contig of 1613 bp in length
201637	201656: gap of unknown length
201657	202403: contig of 747 bp in length



Quality coverage: agarose-PP - N/A  
Quality coverage: 5.2x sum-of-coverage - N/A

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 22315: contig of 22315 bp in length  
\* 22316: gap of unknown length  
\* 22316: contig of 1861 bp in length  
\* 41197: gap of unknown length  
\* 41197: contig of 11387 bp in length  
\* 52604: gap of unknown length  
\* 52604: contig of 17734 bp in length  
\* 70358: gap of unknown length  
\* 70358: contig of 11859 bp in length  
\* 82237: gap of unknown length  
\* 82237: contig of 9686 bp in length  
\* 91943: gap of unknown length  
\* 91943: contig of 9960 bp in length  
\* 101923: gap of unknown length  
\* 101923: contig of 10058 bp in length  
\* 112001: gap of unknown length  
\* 112001: contig of 11906 bp in length  
\* 123927: gap of unknown length  
\* 123927: contig of 9016 bp in length  
\* 132963: gap of unknown length  
\* 132963: contig of 10126 bp in length  
\* 143109: gap of unknown length  
\* 143109: contig of 7062 bp in length  
\* 150191: gap of unknown length  
\* 150191: contig of 6072 bp in length  
\* 156282: gap of unknown length  
\* 156282: contig of 7780 bp in length  
\* 164083: gap of unknown length  
\* 164083: contig of 5200 bp in length  
\* 169303: gap of unknown length  
\* 169303: contig of 7566 bp in length  
\* 176889: gap of unknown length  
\* 176889: contig of 4800 bp in length  
\* 181709: gap of unknown length  
\* 181709: contig of 3420 bp in length  
\* 187149: gap of unknown length  
\* 187149: contig of 6939 bp in length  
\* 194108: gap of unknown length  
\* 194108: contig of 5031 bp in length  
\* 199158: gap of unknown length  
\* 199158: contig of 2330 bp in length  
\* 201509: gap of unknown length  
\* 201509: contig of 2820 bp in length  
\* 204369: gap of unknown length  
\* 204369: contig of 2606 bp in length  
\* 206995: gap of unknown length  
\* 206995: contig of 2753 bp in length  
\* 209767: gap of unknown length  
\* 209767: contig of 3120 bp in length  
\* 212888: gap of unknown length  
\* 212888: contig of 2888 bp in length  
\* 212908: gap of unknown length  
\* 212908: contig of 2888 bp in length.

## FEATURES

Location/Qualifiers  
1. 215795

/organism="Mus musculus"

/strain="C57BL6/J"

/db\_xref="taxon:10090"

/clone="RP23-337K7"

/sex="male"

BASE COUNT 59598 a 48413 c 47378 g 59789 t 617 others  
ORIGIN

Query Match 5.0%; Score 50; DB 55; Length 215795;  
Best Local Similarity 100.0%; Pred. No. 2.9e-14;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

269 acagggatcggatccatcacatggtgtgagccaccatggtttc 318  
Db 129203 ACAGGCATGAGATCCATCACATGAGTGTGAGCCACCACATGTTGTTTC 129154

## RESULT 5

AC025910

LOCUS

DEFINITION

AC025910.3

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC025910 39751 bp DNA HTG 15-JUN-2000

Mus musculus chromosome 11 clone RP23-4004, \*\*\* SEQUENCING IN

PROGRAMS \*\*\*, 26 unordered pieces.

AC025910.3 GI:8248607

HTG: HTGS\_PHASE1.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 39751)

Wetzel, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,

Dederich, D., Thomas, S., Okunou, G., Carlock, C., Garner, T.,

Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,

Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,

Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,

Fernandez, C., Ferraguto, J., Forcum-Tansey, J., Gill, R.,

Corrall, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hognes, M.,

Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,

Kovar, C., Liu, J., Liu, W., Lohse, H., Lozano, R., Martin, R.,

Massey, E., McLeod, M.P., Mel, G., Moore, S., Morgan, M., Morris, S.,

Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,

Perez, L., Renter, D., Say, J., Shen, H., Vasquez, L., Wallington, S.,

Williamson, A., Wrenford, G., Zhou, X., Bouck, J., Hodgson, A.,

Murray, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,

Worley, K., and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 39751)

Worley, K.C.

Direct Submission

Submitted (17-MAR-2000) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jun 5, 2000 this sequence version replaced gi:7264171.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: MACK

Center clone name: RP23-4004

Summary Statistics

Sequencing vector: M13: L08821

Chemistry: Dye-primer Bodipy: 99% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 26845 bases at least Q40

Consensus quality: 42495 bases at least Q30

Consensus quality: 48657 bases at least Q20

Estimated insert size: 26936; sum-of-coverage estimation

Quality coverage: 0.2x in Q20 bases; sum-of-coverage estimation

\*\*\*\*\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3541: contig of 3541 bp in length

```

* 3542 3641: gap of unknown length
* 3642 5191: contig of 1550 bp in length
* 5192 5291: gap of unknown length
* 5292 7282: contig of 1991 bp in length
* 7283 7382: gap of unknown length
* 7383 8811: contig of 1429 bp in length
* 8812 8912: gap of unknown length
* 8912 10410: contig of 1499 bp in length
* 10411 10510: gap of unknown length
* 10511 12264: contig of 1754 bp in length
* 12265 12364: gap of unknown length
* 12365 13735: contig of 1371 bp in length
* 13736 13835: gap of unknown length
* 13836 15189: contig of 1354 bp in length
* 15190 15289: gap of unknown length
* 15290 16602: contig of 1313 bp in length
* 16603 16703: gap of unknown length
* 16703 18483: contig of 1781 bp in length
* 18484 19751: contig of 1168 bp in length
* 19752 21336: contig of 1485 bp in length
* 21337 21436: gap of unknown length
* 21437 22724: contig of 1288 bp in length
* 22725 22824: gap of unknown length
* 22825 23867: contig of 1043 bp in length
* 23868 23967: gap of unknown length
* 23968 25125: contig of 1158 bp in length
* 25126 25225: gap of unknown length
* 25226 26661: contig of 1336 bp in length
* 26662 27997: gap of unknown length
* 27998 28097: contig of 1336 bp in length
* 28098 29123: contig of 1026 bp in length
* 29124 29223: gap of unknown length
* 29224 30371: contig of 1148 bp in length
* 30372 30471: gap of unknown length
* 30472 31715: contig of 1244 bp in length
* 31716 31815: gap of unknown length
* 31816 33136: contig of 1321 bp in length
* 33137 33237: gap of unknown length
* 33238 34241: contig of 1005 bp in length
* 34242 34341: gap of unknown length
* 34342 35698: contig of 1357 bp in length
* 35699 35798: gap of unknown length
* 35799 36753: contig of 1177 bp in length
* 36754 37073: gap of unknown length
* 37074 38458: contig of 1383 bp in length
* 38459 38558: gap of unknown length
* 38559 39751: contig of 1193 bp in length.

```

FEATURES

SOURCE

BASE COUNT 9754 a 8870 c 8644 g 9960 t 2523 others

ORIGIN

Query Match 4.9%; Score 49; DB 50; Length 39751;  
Best Local Similarity 100.0%; Pred. No. 1.1e-13;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 350 ggaattgagccagagccttggagagcagtcgctcttaacctct 398  
|  
Db 17532 GGAATTGAGCCAGAGCCTTGGAGAGCAGTCTCTTAACCTCT 17580

RESULT 6  
AP000028/c  
ID AP000028 standard; DNA: HMW; 45966 BP.

XX

```

AC AP000028;
XX
SV AP000028.1
XX
DI 13-MAY-1998 (Rel. 55, Created)
DT 21-NOV-1999 (Rel. 61, last updated, Version 4)
XX
DE Homo sapiens genomic DNA, chromosome 21q22.1, segment 1/2.
XX
KW HTG.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
XX
RA Hattori M., Ishii K., Toyoda A., Shiba T., Sakaki Y.;
RP 1-45966
[1]
Submitted (11-MAY-1998) to the EMBL/GenBank/DBJ databases.
RL Masahira Hattori, Kitasato University, Department of Science, JST
Sequencing Laboratory, Kitasato 1-15-1, Sagamihara 228, Japan
RL (E-mail: hattori@hgc.ims.u-tokyo.ac.jp, Tel: 0427-78-9732, Fax: 0427-78-9561)
XX
XX [2]
RP 1-45966
RA Hattori M., Ishii K., Toyoda A., Shiba T., Sakaki Y.;
RT "Homo sapiens genomic DNA, chromosome 21q";
RL Unpublished.
XX
CC This sequence is conducted by Kitasato University JST sequencing
laboratory as a JST sequencing team.
CC Principal Investigator: Yoshiyuki Sakaki Ph.D.
CC Phone: +81-3-5449-5622, Fax: +81-3-5449-5445,
CC sakaki@hgc.ims.u-tokyo.ac.jp
CC Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D.
CC The sequence is submitted by Human Genome Sequencing in ALIS
CC project of JST
CC Japan Science and Technology Corporation (JST)
CC 5-3, Yondancho, Chiyoda-ku, Tokyo 102-0028 Japan
CC For further information about this sequence, including its
CC location and relationship to other sequences, please visit our
CC sequence archive Web site (http://www.alls.tokyo.jst.go.jp/HGS/top.html)
CC or send email to webmaster@www.alls.tokyo.jst.go.jp";
XX
FH Key
FH Location/Qualifiers
FT source
FT 1..45966
FT :/chromosome="21"
FT :/db_xref="taxon:9606"
FT :/sequenced_mol="DNA"
FT :/organism="Homo sapiens"
FT :/map="21q22.1"
XX
SQ Sequence 45966 BP; 14010 A; 8695 C; 9324 G; 13937 T; 0 other;

```

Query Match 4.9%; Score 49; DB 14; Length 45966;  
Best Local Similarity 100.0%; Pred. No. 1e-13;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 350 ggaattgagccagagccttggagagcagtcgctcttaacctct 398  
|  
Db 23041 GGAATTGAGCCAGAGCCTTGGAGAGCAGTCTCTTAACCTCT 22993

RESULT 7  
AC025964/c  
LOCUS AC025964 201251 bp DNA HTG 09-JUL-2000  
DEFINITION Mus musculus chromosome 11 clone RP23-136c7 map 11, WORKING DRAFT  
SEQUENCE 48 unordered pieces.  
ACCESSION AC025964

VERSION	AC025964.2	GI:9860918
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus	
AUTHORS	1 (bases 1 to 201251)	
TITLE	Bliren, B., Linton, L., Nusbaum, C. and Lander, E.	
JOURNAL	Mus musculus chromosome 11, clone RP23-156C7	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 201251)	
	Bliren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,	
	Baldwin, J., Batra, N., Bastien, V., Bede, F.,	

TITLE Direct Submission  
 JOURNAL Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 9, 2000 this sequence version replaced gi:726752.  
 COMMENT All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center -----

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*      1         1487: contig of 1487 bp in length
*
*      1488 1587: gap of          100 bp
*
*      1588 2793: contig of 1206 bp in length
*
*      2794 2893: gap of          100 bp
*
*      2894 4247: contig of 1354 bp in length
*
*      4248 4347: gap of          100 bp

```

4348	5649:	contlg	of	1502	bp	in	length
5550	5949:	gap	of	100	bp		
5650	7448:	contlg	of	1299	bp	in	length
7249	7348:	gap	of	100	bp		
7349	8876:	contlg	of	1528	bp	in	length
8677	8976:	gap	of	100	bp		
8677	10439:	contlg	of	1463	bp	in	length
10440	10539:	gap	of	100	bp		
10540	12080:	contlg	of	1541	bp	in	length
12081	12180:	gap	of	100	bp		
12181	14702:	contlg	of	2522	bp	in	length
14703	14802:	gap	of	100	bp		
14803	16733:	contlg	of	1931	bp	in	length
16734	16833:	gap	of	100	bp		
16834	18755:	contlg	of	1922	bp	in	length
18756	18835:	gap	of	100	bp		
18836	20788:	contlg	of	1931	bp	in	length
20787	20886:	gap	of	100	bp		
20887	23125:	contlg	of	2239	bp	in	length
23216	23225:	gap	of	100	bp		
23216	25267:	contlg	of	2042	bp	in	length
25268	25367:	gap	of	100	bp		
25368	27294:	contlg	of	1927	bp	in	length
27295	27396:	gap	of	100	bp		
27395	29083:	contlg	of	1689	bp	in	length
29084	29183:	gap	of	100	bp		
29184	33810:	contlg	of	2627	bp	in	length
31811	31910:	gap	of	100	bp		
31911	334614:	contlg	of	2704	bp	in	length
34615	34714:	gap	of	100	bp		
34715	36824:	contlg	of	2110	bp	in	length
36825	36924:	gap	of	100	bp		
36925	39358:	contlg	of	2434	bp	in	length
39359	39458:	gap	of	100	bp		
39459	41675:	contlg	of	2217	bp	in	length
41676	41775:	gap	of	100	bp		
41776	44442:	contlg	of	2367	bp	in	length
44143	44242:	gap	of	100	bp		
44243	45315:	contlg	of	1133	bp	in	length
45376	45475:	gap	of	100	bp		
45476	48030:	contlg	of	2555	bp	in	length
48031	48130:	gap	of	100	bp		
48131	50762:	contlg	of	2632	bp	in	length
50763	50862:	gap	of	100	bp		
50863	54865:	contlg	of	4003	bp	in	length
54866	54965:	gap	of	100	bp		
54966	58766:	contlg	of	3821	bp	in	length
58767	58886:	gap	of	100	bp		
58887	62079:	contlg	of	3193	bp	in	length
62080	62179:	gap	of	100	bp		
62180	66766:	contlg	of	4585	bp	in	length
66769	66868:	gap	of	100	bp		
66869	70876:	contlg	of	4008	bp	in	length
70877	70976:	gap	of	100	bp		
70977	75996:	contlg	of	5020	bp	in	length
75997	76096:	gap	of	100	bp		
76097	80273:	contlg	of	4177	bp	in	length
80274	80373:	gap	of	100	bp		
80374	85679:	contlg	of	5306	bp	in	length
85680	85779:	gap	of	100	bp		
85780	91711:	contlg	of	5938	bp	in	length
91718	91817:	gap	of	100	bp		
91818	97046:	contlg	of	5229	bp	in	length
97047	97146:	gap	of	100	bp		
97147	101127:	contlg	of	3981	bp	in	length
101128	101227:	gap	of	100	bp		
101228							

```
misc_feature 1.1467000000000002 "assembly-fragment"
misc_feature 1588.2793 "note=assembly-fragment"
misc_feature 2894.4247 "note=assembly-fragment"
misc_feature 4348.5849 "note=assembly-fragment"
misc_feature 5950.7248 "note=assembly-fragment"
misc_feature 7345.8876 "note=assembly-fragment"
misc_feature 8977.10439 "note=assembly-fragment"
misc_feature 10540.12080 "note=assembly-fragment"
misc_feature 12181.14702 "note=assembly-fragment"
misc_feature 14803.16733 "note=assembly-fragment"
misc_feature 16834.18755 "note=assembly-fragment"
misc_feature 18856.20786 "note=assembly-fragment"
misc_feature 20887.23125 "note=assembly-fragment"
```

```
Query Match      4.9%; Score 49; DB 50; Length 201251;
Best Local Similarity 100.0%; Pred. No. 9.3e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

QY 350 ggaattgacctcaggacctttggaagagcagtcagtgctcttaacctc 398  
|||||  
Db 190864 GGAATTGAACTCAGGACCTTGGAGAGACAGTCAGTGCCTTAACCTCT 190816  
|||||

RESULT	8
AC023934/c	
LOCUS	
DEFINITION	AC023934 207757 bp DNA HTG 16-MAR-2000 Homo sapiens chromosome 17 clone RP11-618P13 map 17, WORKING DRAFT
ACCSSION	SOURCE: 35 unordered pieces.
VERSION	AC023934
KEYWORDS	AC023934.2 GI:7249213
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT. human.
ORGANISM	Homo sapiens

**REFERENCE**

1 (bases 1 to 207757)

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS  
 TITLE  
 Homo sapiens chromosome 17, clone RP11-618P13  
 JOURNAL  
 Unpublished  
 2 (bases 1 to 207757)  
 REFERENCE  
 AUTHORS  
 Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.

**TITLE** Direct Submission  
**JOURNAL** Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Mar 15, 2000 this sequence version replaced 417000905

Smilt, A.E.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 -----Genome Center

Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
 Project Information  
 Genar Product Page  
 Genar Institute/ MIT Center for Genome Research  
 Genar code: WMR

```

----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; alignment 0.960731
Consensus quality: 18504 bases at least Q40
Consensus quality: 16210 bases at least Q30
Consensus quality: 200872 bases at least Q20
Insert size: 157000; agarose-gel

```

Quantity coverage: 4.4 In Q20 bases; agarose-fp  
 Quality coverage: 3.3 In Q20 bases; sum-of-contigs

\* consists of 35 collages. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the collages are represented as  
\* runs of 'N', but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1      1178: contig of 1178 bp in length
*      1179: gap of 100 bp
*      1278: contig of 153 bp in length
1379   1431: contig of 100 bp
1379   1431: contig of 100 bp
1532   1531: gap of 100 bp
1532   2672: contig of 1141 bp in length
*      2773: gap of 100 bp
*      2773: contig of 1081 bp in length
*      3853: contig of 100 bp
*      3854: gap of 100 bp
*      3954: gap of 100 bp
*      3954: contig of 1541 bp in length
*      5495: gap of 100 bp
*      5495: gap of 100 bp
*      5595: gap of 1214 bp in length
*      6809: contig of 100 bp
*      6809: gap of 100 bp
*      6909: contig of 1913 bp in length

```

8822	8921:	gap of	100	bp	in	length
8922	10327:	contig of	1306	bp	in	length
10228	10327:	gap of	100	bp	in	length
10328	12417:	contig of	2090	bp	in	length
12418	12517:	gap of	100	bp	in	length
12518	14413:	contig of	1886	bp	in	length
14414	14513:	gap of	100	bp	in	length
14514	16780:	contig of	2267	bp	in	length
16781	16880:	gap of	100	bp	in	length
16881	19899:	contig of	2109	bp	in	length
18990	19089:	gap of	100	bp	in	length
19090	21339:	contig of	2250	bp	in	length
21340	21439:	gap of	100	bp	in	length
21440	24496:	contig of	3057	bp	in	length
24497	24596:	gap of	100	bp	in	length
24597	27488:	contig of	2892	bp	in	length
27489	27588:	gap of	100	bp	in	length
27589	30111:	contig of	2533	bp	in	length
30112	30211:	gap of	100	bp	in	length
30212	33587:	contig of	3376	bp	in	length
33588	33687:	gap of	100	bp	in	length
33688	37578:	contig of	3891	bp	in	length
37579	37678:	gap of	100	bp	in	length
37679	42795:	contig of	5117	bp	in	length
42796	42895:	gap of	100	bp	in	length
42896	48285:	contig of	5390	bp	in	length
48286	48385:	gap of	100	bp	in	length
48386	52171:	contig of	3786	bp	in	length
52172	52271:	gap of	100	bp	in	length
52272	60414:	contig of	8143	bp	in	length
60415	60514:	gap of	100	bp	in	length
60515	66388:	contig of	7872	bp	in	length
66387	68486:	gap of	100	bp	in	length
68487	75220:	contig of	6734	bp	in	length
75221	75320:	gap of	100	bp	in	length
75321	85520:	contig of	7200	bp	in	length
85521	82620:	gap of	100	bp	in	length
82621	92012:	contig of	9392	bp	in	length
92013	92112:	gap of	100	bp	in	length
92113	100583:	contig of	8473	bp	in	length
100584	100685:	gap of	100	bp	in	length
100686	111797:	contig of	11112	bp	in	length
111798	111897:	gap of	100	bp	in	length
111898	120279:	contig of	8382	bp	in	length
120280	120379:	gap of	100	bp	in	length
120380	132336:	contig of	12057	bp	in	length
132337	132336:	gap of	100	bp	in	length
132337	144076:	contig of	11540	bp	in	length
144077	144176:	gap of	100	bp	in	length
144177	160302:	contig of	16126	bp	in	length
160303	160403:	gap of	100	bp	in	length
160404	175068:	contig of	14666	bp	in	length
175069	175168:	gap of	100	bp	in	length
175169	192239:	contig of	17071	bp	in	length
192240	193359:	gap of	100	bp	in	length
193360	207733:	contig of	13418	bp	in	length

```
misc_feature /note="assembly_fragment"  
3954. 5494  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
5595. 6808  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
6909. 8821  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
8922. 10227  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
10328. 12417  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
12518. 14413  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
14514. 16780  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
16881. 18989  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
19090. 21339  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
21440. 24496  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
24597. 27468  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
27589. 30111  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
30212. 33587  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
33688. 37578  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
37679. 42795  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
42896. 48285  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
48386. 52171  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
52272. 60414  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
60515. 68386  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
68487. 75220  
/note="assembly_fragment"  
misc_feature /note="assembly_fragment"  
75321. 82520  
/note="assembly_fragment"
```

Query Match	4.98;	Score 49;	DB 49;	Length 207757;
Best Local Similarity	100.08;	Pred. No. 9.3e-14;		
Matches	49;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;

QY	350	ggatttgaac	ctcagagac	cttcttggaagagacg	atcgtgtctcttaac	ccctc	398
Db	8083	GGAAATGAAC	CTCAGACCTT	TGGAGAGAGCA	GTCACTGCTTAA	CTCT	8035
RESULT	9	AC073817	267762	bp	DNA	HTG	18-JUL-2000
LOCUS	AC073817/c						
DEFINITION		Mus musculus	clone R23-70E17,	WORKING DRAFT	SEQUENCE, 12	ordered	
ACCESSION		AC073817					
VERSION		AC073817.2	GI:9256814				
KEYWORDS		HTG; HTGS_PHASE2; HTGS_DRAFT.					
SOURCE		house mouse.					
ORGANISM		Mus musculus					
REFERENCE		Emmaliot; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS		Eukaryote; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
TITLE		1 (bases 1 to 267762)					
JOURNAL		DOE Joint Genome Institute.					
REFERENCE		2 (bases 1 to 267762)					
AUTHORS		DOE Joint Genome Institute.					
TITLE		Direct Submission					
JOURNAL		Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint					

## COMMENT

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Jul 18, 2000 this sequence version replaced g1:8810434.

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

-----Project Information  
Center Project Name: 1764129  
Center clone name: RPCI-23\_70P17

-----Summary Statistics  
Consensus quality: 254235 bases at least Q40  
Consensus quality: 264194 bases at least Q30  
Consensus quality: 265817 bases at least Q20  
Estimated insert size: 218000; agarose-IP estimation  
Quality coverage: 7.88 in Q20 bases; agarose-IP estimation  
Quality coverage: 6.43 in Q20 bases; sum-of-contigs estimation  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 44594: contig of 44594 bp in length  
\* 44595 44594: gap of unknown length  
\* 44695 44594: contig of 11142 bp in length  
\* 55836 44594: gap of unknown length  
\* 55937 123412: contig of 67476 bp in length  
\* 123413 123412: gap of unknown length  
\* 123512 123412: contig of 61496 bp in length  
\* 185108 123512: gap of unknown length  
\* 185109 185108: contig of 8529 bp in length  
\* 193638 185109: gap of unknown length  
\* 193737 193638: contig of 5569 bp in length  
\* 199307 193737: gap of unknown length  
\* 199406 199307: gap of unknown length  
\* 203254 199406: contig of 3848 bp in length  
\* 203355 203254: gap of unknown length  
\* 207274 203355: contig of 3920 bp in length  
\* 207375 207274: gap of unknown length  
\* 235541 207375: contig of 28166 bp in length  
\* 235641 235541: gap of unknown length  
\* 243223 235641: contig of 7583 bp in length  
\* 243324 243223: gap of unknown length  
\* 243324 243323: contig of 23299 bp in length  
\* 266623 243324: gap of unknown length  
\* 266723 266623: contig of 1040 bp in length.  
\* 267762 266723: gap of unknown length  
\* 267762 267762: contig of 1040 bp in length.  
\* Location/Qualifiers  
1. 267762  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="RPCI mouse BAC library 23"  
ORIGIN  
79382 a 57919 c 57007 g 72350 t 1104 others

Base Count 79382 a 57919 c 57007 g 72350 t 1104 others  
ORIGIN  
Query Match 4.98; Score 49; DB 57; Length 267762;  
Best Local Similarity 100.0%; Pred. No. 9.1e-14;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 ggaattgaactcagagccttgagagagcgtatgctcttaactctc 398  
|||||  
Db 62663 GGAATTGAACCTCAGGACCTTTGAGAGGACGACGTCCTTAACCTCT 62615  
|||||  
RESULT 10  
AC026760  
LOCUS AC026760 212481 bp DNA HTG 21-JUL-2000

## DEFINITION

Mus musculus clone RPCI-80P11 strain C57BL/6/J, WORKING DRAFT  
SEQUENCE: 30 unordered pieces.

ACCESSION AC026760.8 GI:9309493  
VERSION AC026760.8  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## REFERENCE

Submitted (24-MAR-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA  
On Jul 21, 2000 this sequence version replaced g1:8389339.  
-----Genome Center Center:  
Albert Einstein College of Medicine Center  
Code: AECOM  
Web site:  
http://sequence.aecom.yu.edu/cgi-  
bin/vs\_exe/mousedb/mouseseq/mouseseqtable.htm  
Contact: jhanssequence.aecom.yu.edu

-----Summary Statistics  
Center project name: ABU  
Sequencing vector: pUC18; L08752  
Chemistry: Dye-terminator Big Dye; 100%  
\*Consensus quality: 200274 at least Q20  
\*Consensus quality: 194692 at least Q30  
\*Consensus quality: 184737 at least Q40  
\*Estimated insert size: agarose-IP - N/A  
\*Estimated insert size: 211901 - sum-of-contigs  
Quality coverage: 5.2x sum-of-contigs - N/A  
Quality coverage: 5.2x sum-of-contigs - N/A

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 30 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 39129: contig of 39129 bp in length  
\* 39130 39129: gap of unknown length  
\* 39130 39129: contig of 16266 bp in length  
\* 55437 39130: gap of unknown length  
\* 55438 55437: contig of 9585 bp in length  
\* 65023 55438: gap of unknown length  
\* 65042 65023: gap of 11061 bp in length  
\* 76103 65042: contig of 11061 bp in length  
\* 76104 76103: gap of unknown length  
\* 76123 76104: gap of unknown length  
\* 89530 76123: contig of 13407 bp in length  
\* 89531 89530: gap of unknown length  
\* 102268 89531: contig of 12718 bp in length  
\* 102269 102268: gap of unknown length  
\* 112684 102269: contig of 10396 bp in length  
\* 112685 112684: gap of unknown length  
\* 112705 112685: contig of 10750 bp in length  
\* 123445 112705: gap of unknown length  
\* 123475 123445: gap of unknown length  
\* 123475 123474: gap of 6422 bp in length  
\* 129896 123475: contig of 6422 bp in length  
\* 129916 129896: gap of unknown length  
\* 129917 129916: contig of 9926 bp in length  
\* 139842 129917: gap of unknown length  
\* 139862 139842: gap of unknown length  
\* 147484 139862: contig of 7622 bp in length  
\* 147485 147484: gap of unknown length

147505 154936: contig of 7432 bp in length  
 \* 154937 154936: gap of unknown length  
 \* 154957 161276: contig of 6320 bp in length  
 \* 161277 161296: gap of unknown length  
 \* 161297 166440: contig of 5144 bp in length  
 \* 166441 166460: gap of unknown length  
 \* 166461 166648: contig of 188 bp in length  
 \* 166649 166668: gap of unknown length  
 \* 166669 170718: contig of 4050 bp in length  
 \* 170719 170738: gap of unknown length  
 \* 170739 175460: contig of 4722 bp in length  
 \* 175461 175480: gap of unknown length  
 \* 175481 179447: contig of 3967 bp in length  
 \* 179448 182689: contig of 3232 bp in length  
 \* 179468 182719: gap of unknown length  
 \* 182720 188035: contig of 5316 bp in length  
 \* 188036 188055: gap of unknown length  
 \* 188056 190268: contig of 2213 bp in length  
 \* 190269 190288: gap of unknown length  
 \* 190289 194882: contig of 4594 bp in length  
 \* 194883 194902: gap of unknown length  
 \* 194903 197472: contig of 2550 bp in length  
 \* 197473 197473: gap of unknown length  
 \* 197474 200813: contig of 3341 bp in length  
 \* 200814 204810: contig of 3977 bp in length  
 \* 204811 204830: gap of unknown length  
 \* 204831 205092: contig of 262 bp in length  
 \* 205093 205112: gap of unknown length  
 \* 205113 207503: contig of 2391 bp in length  
 \* 207504 207523: gap of unknown length  
 \* 207524 209358: contig of 1835 bp in length  
 \* 209359 209378: gap of unknown length  
 \* 209379 211271: contig of 1893 bp in length  
 \* 211272 211292: gap of unknown length  
 \* 211292 212481: contig of 1190 bp in length.

FEATURES  
 source  
 1. 212481  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RP23-80p11"  
 /sex="male"

BASE COUNT  
 55156 a 50971 c 51342 g 54366 t 646 others

Query Match 4.8%; Score 48; DB 51; Length 212481;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-13;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 348 caggaatgacacagcagccttggaagagcagctgctcttaacc 395  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 8839 CAGGAATTGAACTCAGACCTTGGAAGAGCACTGCTCTTAACC 8886

RESULT 11  
 AF168591 118 bp DNA ROD 29-SEP-1999  
 LOCUS Rattus norvegicus clone C4 repetitive sequence B2.  
 DEFINITION AF168591  
 ACCESSION AF168591.1 GI:5739358  
 VERSION  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 118)  
 Liu, X., Clemens, J.A., Yin, T., Stephenson, D.T., Johnstone, E.M.,  
 Du, Y., Panetta, J.A., Paul, S.M. and Little, S.P.  
 TITLE Rattus B(2) sequences are induced in the hippocampal CA1 region after

JOURNAL transient global cerebral ischemia  
 MEDLINE U. Biol. Chem. 274 (40), 28674-28681 (1999)  
 REFERENCE 99428547;  
 2 (bases 1 to 118)  
 AUTHORS Liu, X. and Little, S.P.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-JUN-1999) Neuroscience Research Division, E11 L111Y  
 and Company, 355 East Merrill Street, Indianapolis, IN 46285, USA  
 Location/Qualifiers

FEATURES

source  
 1. 118  
 /organism="Rattus norvegicus"  
 /strain="Wistar"  
 /db\_xref="taxon:10116"  
 /clone="C4"  
 /tissue\_type="hippocampus"

repeat\_region

1. 118  
 /rpt\_family="B2"  
 /rpt\_type="dispersed"

BASE COUNT 23 a 33 c 30 g 32 t

Query Match 4.6%; Score 46; DB 88; Length 118;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-12;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 271 aggcacgcacccattacagatggtgagccacacatgtggtt 316  
 ||||||||||||||||||||||||||||||||||||||||||||  
 Db 105 AGGCATCGCATCCATTACAGATGTTGAGCCACCATGTGTT 60

RESULT 12  
 RN012268/c 1201 bp mRNA ROD 26-OCT-1994  
 LOCUS RN012268/c  
 DEFINITION Rattus norvegicus carbonic anhydrase V mRNA, complete cds.  
 ACCESSION U12268  
 VERSION U12268.1 GI:517476  
 KEYWORDS mitochondrial; carbonic anhydrase V.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 1201)  
 Nagao, Y., Svinivashan, M., Platero, J.S., Svendrowaki, M., Wahed, A.  
 and Sly, W.S.  
 TITLE Mitochondrial carbonic anhydrase (isozyme V) in mouse and rat: CDNA  
 cloning, expression, subcellular localization, processing, and  
 tissue distribution  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91, 10330-10334 (1994)

MEDLINE 95024107;  
 REFERENCE 2 (bases 1 to 1201)  
 Nagao, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-JUN-1994) Yoshiro Nagao, Department of Biochemistry,  
 Saint Louis University, 1402 South Grand Boulevard, St. Louis, MO  
 63104, USA  
 Location/Qualifiers

FEATURES  
 source  
 1. 1201  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="11b+lambda ZAPII rat liver, StrataGene catalog No.  
 936513"

sig\_peptide  
 CDS  
 16..117  
 /evidence="experimental"  
 16..930  
 /EC\_number="4.2.1.1"  
 /note="mitochondria matrix"  
 /codon\_start=1  
 /product="carbonic anhydrase V"  
 /protein\_id="AAA50832.1"

```
/db_xref="GI:522180"
/translation="MRAKMLGSGPYKPLAIRHNGPLCATRPPQMRQSHYAERHSHN
CARHPLMTGTPVSPGCTQSPINQIOTDSVDPKLAFLRVSYDAASCRLMNTGTFQ
VERDDSCSESGISGGLGNHRYLKOFHFMGATDEGSEHMDGHAAPAEHLVHNS
MYENIKKATGSENGLAVIGVFLKGAHHAALQRLVDILPEVRHKDOTVMGPPRSC
LIPACRDYWTYPSLTTPPLASVTWIVHKMPLLEVSPQSLSTFRLLTSGRDEDEVA
VNFRLQPLRLGRNRNVSFQVPRVGRKS"
polysignal 1185.1190
BASE COUNT 279 a 340 c 330 g 252 t
ORIGIN

Query Match 4.6%; Score 46; DB 88; Length 1201;
Best Local Similarity 100.0%; Pred. No. 4.4e-12;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 agggcgcgtccatccatcagatggtgtgagccacatggtgt 316
|||||
Db 1135 AGGCGATCGGATCCCATTCACAGTGTGTGAGCCACCATGTGTT 1090

RESULT 13
AF095690 3988 bp DNA ROD 04-FEB-1999
LOCUS Mus musculus fibronectin (Fn) gene, alternatively spliced products,
DEFINITION partial cds.
ACCESSION AF095690
VERSION AF095690.1 GI:4218964
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3988)
AUTHORS Muro,A.F., Iaconicig,A. and Baralle,F.E.
TITLE Regulation of the fibronectin EDA exon alternative splicing.
FEATURES
JOURNAL FEBS Lett. 437 (1-2), 137-141 (1998)
MEDLINE 99019332
REFERENCE 2 (bases 1 to 3988)
AUTHORS Muro,A.F., Iaconicig,A. and Baralle,F.E.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-1998) Pathology, ICGEB, Padriciano 99, Trieste,
TS 34012, Italy
FEATURES
source location/Qualifiers
|>.3988
/organism="Mus musculus"
/strain="129SV"
/db_xref="taxon:10090"
<1..>3988
/gene="Fn"
join(<184..297,2772..>2959)
/gene="Fn"
/note="EDA (EIIIA or EDI) alternatively spliced exon 1s
excluded"
/product="fibronectin"
join(<184..297,1332..1601,2772..>2959)
/gene="Fn"
/note="EDA (EIIIA or EDI) alternatively spliced exon 1s
included"
/product="fibronectin"
184..297
/gene="Fn"
join(<184..297,1332..1601,2772..>2959)
/gene="Fn"
/note="EDA (EIIIA or EDI) alternatively spliced exon 1s
included"
/codon_start=3
/product="fibronectin"
/protein_id="AADI2250.1"
/db_xref="GI:4218965"
/translation="QTEKTIKGLQPTVEYVSVTAONRNGESPLVQTAVTNIDRPGK

polysignal 1185.1190
BASE COUNT 279 a 340 c 330 g 252 t
ORIGIN

Query Match 4.6%; Score 46; DB 88; Length 1201;
Best Local Similarity 100.0%; Pred. No. 4.4e-12;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 agggcgcgtccatccatcagatggtgtgagccacatggtgt 316
|||||
Db 1135 AGGCGATCGGATCCCATTCACAGTGTGTGAGCCACCATGTGTT 1090

RESULT 14
MMMH04 10000 bp DNA ROD 13-MAR-1994
LOCUS Mouse gene coding for major histocompatibility antigen. This is a
DEFINITION class II antigen, I-A-beta.
ACCESSION V01527
VERSION V01527.1 GI:53081
KEYWORDS histocompatibility antigen; signal peptide.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 10000)
AUTHORS Larhammar,D., Hammerling,U., Denaro,M., Lund,T., Flavell,R.A.,
Rask,L. and Peterson,P.A.
TITLE Structure of the murine immune response I-A beta locus: sequence of
the I-A beta gene and an adjacent beta-chain second domain exon
JOURNAL Cell 34 (1), 179-188 (1983)
MEDLINE 83285340
REFERENCE 2 (bases 1 to 10000)
AUTHORS Larhammar,D.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-1983) to the EMBL/GenBank/DBJ databases
COMMENT Data kindly reviewed (12-OCT-1983) by D. Larhammar.
FEATURES
source location/Qualifiers
|>.10000
/organism="Mus musculus"
/db_xref="taxon:10090"
2059..2472
/number=1
exon prim_transcript:2059..>8465
old_sequence /note="primary transcript"
2237..2239
/note="GCT [1] revised to GCT [2]"
/citation=[1]
/citation=[2]
join(2379..2472,3828..4100,6408..6689,6995..7105,
7533..7556,8166..8179)
/gene="I-A-beta"
join(2379..2472,3828..4100,6408..6689,6995..7105,
7533..7556,8166..8179)
/gene="I-A-beta"

polysignal 1185.1190
BASE COUNT 279 a 340 c 330 g 252 t
ORIGIN

Query Match 4.6%; Score 46; DB 88; Length 1201;
Best Local Similarity 100.0%; Pred. No. 4.4e-12;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 agggcgcgtccatccatcagatggtgtgagccacatggtgt 316
|||||
Db 1135 AGGCGATCGGATCCCATTCACAGTGTGTGAGCCACCATGTGTT 1090

RESULT 14
MMMH04 10000 bp DNA ROD 13-MAR-1994
LOCUS Mouse gene coding for major histocompatibility antigen. This is a
DEFINITION class II antigen, I-A-beta.
ACCESSION V01527
VERSION V01527.1 GI:53081
KEYWORDS histocompatibility antigen; signal peptide.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 10000)
AUTHORS Larhammar,D., Hammerling,U., Denaro,M., Lund,T., Flavell,R.A.,
Rask,L. and Peterson,P.A.
TITLE Structure of the murine immune response I-A beta locus: sequence of
the I-A beta gene and an adjacent beta-chain second domain exon
JOURNAL Cell 34 (1), 179-188 (1983)
MEDLINE 83285340
REFERENCE 2 (bases 1 to 10000)
AUTHORS Larhammar,D.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-1983) to the EMBL/GenBank/DBJ databases
COMMENT Data kindly reviewed (12-OCT-1983) by D. Larhammar.
FEATURES
source location/Qualifiers
|>.10000
/organism="Mus musculus"
/db_xref="taxon:10090"
2059..2472
/number=1
exon prim_transcript:2059..>8465
old_sequence /note="primary transcript"
2237..2239
/note="GCT [1] revised to GCT [2]"
/citation=[1]
/citation=[2]
join(2379..2472,3828..4100,6408..6689,6995..7105,
7533..7556,8166..8179)
/gene="I-A-beta"
join(2379..2472,3828..4100,6408..6689,6995..7105,
7533..7556,8166..8179)
/gene="I-A-beta"
```





```

Intron /number-10
10456..11391
exon /number-10
11392..11528
Intron /number-11
11529..11684
exon /number-11
11685..11829
Intron /number-12
11830..11923
exon /number-12
11924..12068
Intron /number-13
12069..13362
exon /number-13
13363..13480
Intron /number-14
13481..13568
exon /number-14
13569..>13574
number-15
BASE COUNT 3292 a 3758 c 3697 g 3302 t
ORIGIN

```

```

Query Match 4.68; Score 46; DB 88; Length 14049;
Best Local Similarity 100.0%; Pred. No. 3.7e-12;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 350 ggaattgaactcagagccttggagaagcagtcgtcttaacc 395
|||||
Db 2835 GGAATTGAACCTCAGGACCTTTGGAAGACGACGTCGCTCTTAACC 2790
|||||

```

Search completed: November 4, 2000, 03:54:39  
 Job time: 20789 sec

us-09-101-423a-3.olig.rng

Mon Nov 6 10:14:17 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:35:34 ; Search time 255.06 Seconds  
(without alignments)  
1484.623 Million cell updates/sec

Title: US-09-101-423A-3  
Perfect score: 1008  
Sequence: 1 aggacagagttcacatccc.....ttcctgtgtcctcctggaac 1008

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 480022 seqs, 187831343 residues

Word size : 0

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	810	80.4	1008	18	Metastasis inducin
2	41	4.1	9050	21	Wild type (C57BL/6
3	39	3.9	13206	11	Rat acyl peptide h
C 4	38	3.8	16956	21	Wild type (C57BL/6
C 5	37	3.7	1210	16	Rat fatty acid bin
C 6	36	3.6	49999	20	Murine LOBO homolo
C 7	35	3.5	4113	12	Mouse U6 Intracul
C 8	33	3.3	5109	17	Rat brain homology
C 9	32	3.2	13206	11	Rat acyl peptide h
10	32	3.2	44576	21	Cosmid CVO14 conta
11	32	3.2	49999	20	Murine LOBO genom
12	32	3.2	49999	20	Murine LOBO homolo

c	13	32	3.2	51259	18	X83007	Partial mouse WRN
	14	31	3.1	26698	17	T08126	Mouse syndecan-1 g
	15	31	3.1	26700	15	O67902	Syndecan gene. Mu
	16	31	3.1	26700	19	V15946	Mouse syndecan gen
	17	31	3.1	26700	20	V81283	Mouse syndecan-1 e
	18	31	3.1	44576	21	Z61522	Cosmid CVO14 conta
c	19	30	3.0	13011	18	T96631	cDNA encoding rat
	20	29	2.9	4283	20	V72672	Hamster 3' DNA seq
	21	29	2.9	16009	20	X86357	Rat smooth muscle
c	22	29	2.9	36901	20	Z23892	Murine LOBO genom
	23	29	2.9	38886	20	Z23897	Murine LOBO homolo
c	24	27	2.7	774	20	X57443	Rat U3 gene trap d
	25	27	2.7	880	20	X57466	Rat U3 gene trap d
c	26	27	2.7	880	20	X57468	Rat U3 gene trap d
c	27	27	2.7	6727	18	T86014	Murine IL-5 cDNA g
	28	27	2.7	13146	18	T96719	Murine RENT1 genom
	29	27	2.7	16009	20	X86357	Rat smooth muscle
	30	27	2.7	2751	15	O66164	Seven transmembran
	31	26	2.6	2751	19	V18351	Murine V31 seven t
	32	26	2.6	2858	16	O75270	Unique 2.9 kb geno
c	33	26	2.6	4243	17	T34501	RHAMM 1 gene intro
	34	26	2.6	7997	16	O85553	Rat cholesterol 7-
	35	26	2.6	7997	16	O85553	Rat cholesterol 7-
	36	26	2.6	2276	16	O87170	Natural resistance
c	37	25	2.5	2276	16	O87170	Natural resistance
c	38	25	2.5	2294	16	O87167	Natural resistance
c	39	25	2.5	2485	16	O87167	Natural resistance
c	40	25	2.5	3519	16	O92937	HIRPE nucleotide s
	41	25	2.5	5006	21	A28965	Mouse A-myb genom
	42	25	2.5	5889	20	V84328	Mouse Friend virus
	43	25	2.5	6480	19	T99572	Mouse Presenilin-1
	44	25	2.5	48974	20	X55300	Wild type (C57BL/6
c	45	25	2.5	90050	21	Z91925	

## ALIGNMENTS

RESULT 1  
T72787  
ID T72787 standard; DNA; 1008 BP.  
XX  
AC T72787;  
XX  
DT 22-SEP-1997 (first entry)  
XX  
DE Metastasis inducing DNA C6.  
XX  
KW Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;  
XX osteopontin; ss,  
XX Homo sapiens.  
XX WO9725443-A1.  
XX 17-JUL-1997.  
XX 10-JAN-1997; 97WO-GB000074.  
XX 10-JAN-1996; 96GB-0000470.  
XX (UPLI-) UNIV LIVERPOOL.  
XX Barracough BR, Rudland PS;  
XX WPI; 1997-372878/34.  
XX New isolated metastasis-inducing DNA - used to develop products to  
XX identify and treat patients at risk from metastatic tumours  
XX Claim 10; Page 26; 38pp; English.  
XX Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12  
CC

and C20 (T72795-90) are entirely novel short stretches of human regulatory DNA capable of inducing metastasis. They were identified using a method for detecting Met-DNA that involves transferring human DNA from a malignant, metastatic cancer cell (in this case, breast cancer) into a cell line (pref. rat Rama 3) that produces only benign, non-metastasizing tumours when injected into a syngeneic animal, injecting those transformed cells into a syngeneic animal, selecting those animals having metastasizing tumours, and recovering the Met-DNA from them. The isolated Met-DNAs can be used to develop products to identify and treat patients at risk from metastatic tumours.

Sequence 1008 BP; 273 A; 228 C; 200 G; 307 T; 0 other;

Query Match 80.4%; Score 810; DB 18; Length 1008;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 860; Conservative 0; Mismatches 1; Reads 0

[illegible]

Db	928	aaaggcagtcctctaaacgctgcgcctcatttcattcatcagagctgaatgcaattctgtgct	987
Qy	988	cttcctgtgctcctcctggaac	1008
Db	988	cttcctgtgctcctcctggaac	1008
RESULT	2		
291925			
ID	291925	standard; DNA; 90050 BP.	
XX			
AC	291925;		
XX			
DT	08-JUN-2000	(first entry)	
XX			
DE	Wild type (C57BL/6J)	murine mahogany protein genomic sequence c110/111.	
XX			
KW	Mahogany gene; mouse; mg gene; regulatory defect; gene therapy; obesity;		
KW	weight regulation; cell therapy; body weight disorder; cachexia;		
KW	anorexia; hyperpigmentation; increased metabolic rate disorder;		
KW	hyperphagia; Anorexia; antianorexic; anticachexia; ds.		
XX			
OS	Mus sp.		
XX			
XX	WO200005373-A2.		
PN			
XX			
PD	03-FEB-2000.		
XX			
PF	21-JUL-1999;	99WO-US16484.	
XX			
PR	21-JUL-1998;	98US-0093630.	
PR	20-OCT-1998;	98US-0104978.	
PR	05-FEB-1999;	99US-0245041.	
XX			
PA	(MILL-) MILLENIUM PHARM INC.		
XX			
PI	Moore K, Nagle DL;		
XX			
DR	WPI; 2000-195103/17.		
XX			
PT	New human and murine mahogany genes, useful, e.g. for diagnosis and		
PT	treatment of body weight disorders -		
XX			
PS	Example; Fig 3d; 188pp; English.		
XX			
CC	This sequence represents a murine mahogany gene of the invention.		
CC	The mahogany genes are used: (i) to produce recombinant mahogany (mg)		
CC	proteins (ii); (ii) as a source of antisense, ribozyme or triplex-forming		
CC	therapeutics; (iii) as a source of diagnostic probes and primers for		
CC	detecting expression of mg genes or mutations, regulatory defects, in		
CC	this gene, or for isolation of related sequences; and		
CC	(iv) in (cell-based) gene therapy. (ii) are used to raise specific		
CC	antibodies (Ab): to identify other (extra)cellular products involved in		
CC	weight regulation, and to screen for agents that disrupt interaction		
CC	between (ii) and other macromolecules. The Ab are used to detect abnormal		
CC	levels (or function) of (ii) for diagnosis, prognosis or monitoring of		
CC	treatment); to evaluate (ii)-expressing cells intended for cell therapy,		
CC	and as therapeutic mg inhibitors. Cells that express the mg gene (or		
CC	contain the mg polypeptide) are used to identify agents (A) that modulate		
CC	mg activity. (A) are potentially useful for the treatment of body weight		
CC	disorders, particularly obesity, cachexia or anorexia, or other		
CC	conditions associated with the mg gene such as hyperpigmentation,		
CC	hyperphagia and disorders that result in increased metabolic rate.		
XX			
XX	Sequence 90050 BP; 25307 A; 18193 C; 19454 G; 27096 T; 0 other:		

```

Query Match      4.1%  Score 41;  DB 21;  Length 90050;
Best Local Similarity 100.0%;  Pred. No. 1.2e-10;
Matches 41;  Conservative 0;  Mismatches 0;  Indels 0;

QY 354 ttgaactcagracctttggaagcagtcagtcagtccttaac 394
      |||||

```

us-09-101-423a-3.olig.rng

Mon Nov 6 10:14:17 2000

Db 73514 ttgaactcaggacotttggaagagcagtcagtgctttaa 73554

RESULT 3

Q05243

ID Q05243 standard; DNA; 13206 BP.

XX AC

Q05243;

XX AC

29-NOV-1990 (first entry)

XX DE Rat acyl peptide hydrolase (APH) gene and flanking regions.

XX KW Acyl peptide hydrolase; APH; cancer; small cell carcinoma; SCC.

XX OS Synthetic.

XX PH Key

FT exon 2580..2842

FT intron /tag= a

FT exon 2843..3072

FT intron /tag= b

FT exon 3073..3216

FT intron /tag= c

FT exon 3217..3290

FT intron /tag= d

FT exon 3291..3423

FT intron /tag= e

FT exon 3424..3875

FT intron /tag= f

FT exon 3876..4002

FT intron /tag= g

FT exon 4003..4748

FT intron /tag= h

FT exon 4749..4842

FT intron /tag= i

FT exon 4843..4934

FT intron /tag= j

FT exon 4935..5010

FT intron /tag= k

FT exon 5011..5099

FT intron /tag= l

FT exon 5100..5263

FT intron /tag= m

FT exon 5264..5404

FT intron /tag= n

FT exon 5405..5542

FT intron /tag= o

FT exon 5542..5636

FT intron /tag= p

FT exon 5637..5728

FT intron /tag= q

FT exon 5729..5804

FT intron /tag= r

FT exon 5805..5845

FT intron /tag= s

FT exon 5846..5936

FT intron /tag= t

FT exon 5937..6058

FT intron /tag= u

FT exon 6059..6520

FT intron /tag= v

FT exon 6521..6581

FT intron /tag= w

FT exon 6582..7584

FT intron /tag= x

FT exon 7585..7682

FT intron /tag= y

FT exon 7683..8283

FT intron /tag= z

FT exon 8284..8335

FT intron /tag= aa

FT exon /tag= ab  
9440..9528  
FT intron /tag= ac  
9529..9886  
FT exon /tag= ad  
9887..10025  
FT intron /tag= ae  
10026..10436  
FT exon /tag= af  
10437..10520  
FT intron /tag= ag  
10521..10668  
FT exon /tag= ah  
10669..10749  
FT intron /tag= ai  
10750..11203  
FT exon /tag= aj  
11204..11292  
FT intron /tag= ak  
11293..11385  
FT exon /tag= al  
11386..11576  
FT intron /tag= am  
11577..11660  
FT exon /tag= an  
11661..11763  
FT intron /tag= ao  
11764..11850  
FT exon /tag= ap  
11851..11957  
FT intron /tag= aq  
11958..12042  
FT exon /tag= ar  
12043..12304  
FT repeat\_unit /tag= as  
1664..1859  
FT repeat\_unit /tag= at  
1860..2047  
FT exon /tag= au  
12043..12304  
XX EP378224-A.  
PN 18-JUL-1990.  
XX 12-JAN-1990; 90EP-0100575.  
XX 13-JAN-1989; 89US-0296996.  
XX 01-NOV-1989; 89US-0429935.  
XX (GEO-) GEN HOSPITAL CORP.  
XX Smith JA;  
XX WPI; 1990-218599/29.  
XX Detecting small cell carcinoma by hybridisation and analysis -  
XX using nucleic acid encoding acyl-peptide hydrolase enzyme.  
XX Disclosure; 47pp; English.  
XX APH may be used in hydrolysis of an N-alpha-acetylated peptide,  
XX probes raised to the peptide product may be used to detect and  
XX identify SCC and other cancers arising from chromosome 3 deletions.  
XX See also Q05241.  
XX Sequence 13206 BP; 3039 A; 3419 C; 3494 G; 3254 T; 0 other;  
SQ

Query Match 3.9%; Score 39; DB 11; Length 13206;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 ggatccattacagatggtgtgagccaccatgtggttt 317

Db 1947 ggatccattacagatggttgagccaccatgtggtt 1985  
 |||

RESULT 4  
 ID 291923 standard; DNA; 16956 BP.  
 XX 291923;  
 AC 291923;  
 DT 08-JUN-2000 (first entry)  
 DE Wild type (C57BL/6J) murine mahogany protein genomic sequence c56.  
 XX Mahogany gene; mouse; mg gene; regulatory defect; gene therapy; obesity;  
 KW weight regulation; cell therapy; body weight disorder; cachexia;  
 KW anorexia; hyperpigmentation; increased metabolic rate disorder;  
 KW hyperphagia; Antiobesity; antianorexic; anticachexic; ds.  
 OS Mus sp.  
 XX WO200005373-A2.  
 PN 03-FEB-2000.  
 XX 21-JUL-1999; 99WO-US16484.  
 XX 21-JUL-1998; 98US-0093630.  
 PR 20-OCT-1998; 98US-0104378.  
 PR 03-FEB-1999; 99US-0245041.  
 XX (MILL-) MILLENIUM PHARM INC.  
 PA Moore K, Nagle DM;  
 PI WPI; 2000-195103/17.  
 XX New human and murine mahogany genes, useful, e.g. for diagnosis and  
 PT treatment of body weight disorders -  
 PT Example; Fig 3b; 188pp; English.

CC This sequence represents a murine mahogany gene of the invention.  
 CC The mahogany genes are used: (i) to produce recombinant mahogany (mg)  
 CC proteins (ii); (ii) as a source of antisense, ribozyme or triplex-forming  
 CC therapeutics; (iii) as a source of diagnostic probes and primers for  
 CC detecting expression of mg genes or mutations, regulatory defects, in  
 CC this gene, or for isolation of related sequences; and  
 CC (iv) in (cell-based) gene therapy. (ii) are used to raise specific  
 CC antibodies (Ab); to identify other (extra)cellular products involved in  
 CC weight regulation, and to screen for agents that disrupt interaction  
 CC between (ii) and other macromolecules. The Ab are used to detect abnormal  
 CC levels (or function) of (ii) (for diagnosis, prognosis or monitoring of  
 CC treatment); to evaluate (ii)-expressing cells intended for cell therapy,  
 CC and as therapeutic mg inhibitors. Cells that express the mg gene (or  
 CC contain the mg polypeptide) are used to identify agents (A) that modulate  
 CC mg activity. (A) are potentially useful for the treatment of body weight  
 CC disorders, particularly obesity, cachexia or anorexia, or other  
 CC conditions associated with the mg gene such as hyperpigmentation,  
 CC hyperphagia and disorders that result in increased metabolic rate.

XX Sequence 16956 BP; 5172 A; 3599 C; 3377 G; 4781 T; 27 other;

Query Match 3.8%; Score 38; DB 21; Length 16956;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-09;  
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 ggatccattacagatggttgagccaccatgtggtt 316  
 |||

Db 15851 GGATCCCATACAGATGTTGTGAGCCACCATGTGTT 15814

RESULT 5  
 ID Q87049 standard; DNA; 1210 BP.  
 XX Q87049;  
 AC Q87049;  
 DT 11-JAN-1996 (first entry)  
 DE Rat fatty acid binding protein (FABP) gene fragment.  
 XX Fatty acid binding protein; promoter; transgene; gastro-intestinal;  
 KW ds.  
 OS Rattus sp.  
 XX WO9511299-A1.  
 PN 27-APR-1995.  
 PD 13-OCT-1994; 94WO-US11716.  
 XX 18-OCT-1993; 93US-0141323.  
 PR (AMGE-) AMGEN INC.  
 PA Ratzkin BJ, Simonet WS;  
 PI WPI; 1995-193793/25.  
 XX Enhanced transgene expression in specific tissues - of the  
 PT gastrointestinal tract of a non-human animal  
 PT Claim 1; Page 33; 52pp; English.

CC Q87049 represents a DNA fragment containing the rat intestinal fatty  
 CC acid binding protein (FABP) promoter, from position 1 and 1210 bp in  
 CC the rat FABP gene. The FABP promoter sequence is used to construct  
 CC an FABP promoter/interleukin-8 transgene which is ultimately  
 CC inserted into a plasmid pRE-IL-8 PA. The plasmid vector is then  
 CC injected into mouse embryos and the transgene is expressed in vivo.  
 CC Depending on the transgene(s) inserted into the animal and the  
 CC level of transgene expression produced, the animal may become  
 CC either increasingly susceptible or less susceptible to certain  
 CC diseases or disease types. A further use for the invention is for  
 CC developing methods useful in diagnosing disease.

XX Sequence 1210 B; 403 A; 200 C; 234 G; 373 T; 0 other;

Query Match 3.7%; Score 37; DB 16; Length 1210;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 280 gatccattacagatggttgagccaccatgtggtt 316  
 |||

Db 317 GATCCCATACAGATGTTGTGAGCCACCATGTGTT 281

RESULT 6  
 ID Z23895 standard; DNA; 49999 BP.  
 XX Z23895;  
 AC Z23895;

DT 25-JAN-2000 (first entry)

DE Murine LOBO homologue genomic DNA fragment 1.

XX LOBO; long bones; bone development; bone extension; skull; osteopathic;  
 KW diagnostic; pharmacological; gene therapy; transgenic animal; disease;  
 KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.  
 OS Mus musculus.

us-09-101-423a-3.olig.rng

Mon Nov 6 10:14:17 2000

```

XX PN W09950284-A2.
XX PD 07-OCT-1999.
XX PF 26-MAR-1999; 99WO-EP02055.
XX PR 27-MAR-1998; 98DE-1013799.
XX PA (ROSE/) ROSENTHAL A.
XX PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX WPI; 1999-601320/51.
XX Nucleic acids encoding proteins which influence bone development,
XX useful for treating and studying bone disorders -
XX Example 3; Page 133-161; 391pp; German.
XX This invention describes novel nucleic acids (I; designated LOBO (long
XX bones)) encoding proteins influencing bone development in mammals. The
XX proteins of the invention reduce and/or inactivate bone extension (i.e.
XX development), with exception of the skull and have osteopathic activity.
XX The nucleic acid molecules, proteins and antibodies can be used in
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX and nucleic acid molecules, etc. are useful for production of transgenic
XX animals, especially a transgenic mouse for the study of diseases
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX achondroplasia. This sequence encodes the murine LOBO protein described
XX in the method of the invention.
XX Sequence 49999 BP; 14550 A; 10214 C; 10719 G; 14516 T; 0 other;
SQ

Query Match 3.6%; Score 36; DB 20; Length 49999;
Best Local Similarity 100.0%; Pred. No. 3.4e-08; Indels 0; Gaps 0;
Matches 36; Conservative 0; Mismatches 0;

QY 279 ggatccattacagatgttgtagccaccatgtgg 314
|||||
DB 1563 GGATCCCATTCAGATGTTGTGAGCCACCATGTGG 1528

RESULT 7
Q13400/C
ID Q13400 standard; DNA: 413 BP.
AC Q13400;
XX
XX 27-NOV-1991 (first entry)
XX Mouse U6 intranuclear low mol. RNA gene.
XX RNA promoter; ss.
XX Mus musculus.
XX JPO3183486-A.
XX 09-AUG-1991.
XX 13-DEC-1989; 89JP-0324444.
XX 13-DEC-1989; 89JP-0324444.
XX (KAGA ) KAGAKU OYOBI KESSEI.
XX WPI; 1991-277587/38.
XX Foreign gene transcription and expression plasmid - in which
XX small nuclear RNA promoter is integrated
XX

PS Claim 4; Page 2; 6pp; Japanese.
XX The DNA represents the mouse U6 intranuclear low mol. RNA gene
XX sequence. Since it is an intranuclear low mol. promoter, it is
XX extremely useful in transcription and prodn. of antisense RNA.
XX See also Q13398 and Q13399.
XX Sequence 413 BF; 135 A; 84 C; 95 G; 99 T; 0 other;
SQ

Query Match 3.5%; Score 35; DB 12; Length 413;
Best Local Similarity 100.0%; Pred. No. 1.1e-07; Indels 0; Gaps 0;
Matches 35; Conservative 0; Mismatches 0;

QY 280 gatccattacagatgttgtagccaccatgtgg 314
|||||
DB 129 GATCCCATTCAGATGTTGTGAGCCACCATGTGG 95

RESULT 8
T72654
ID T72654 standard; cDNA: 5109 BP.
XX
XX T72654;
AC
XX 20-AUG-1997 (first entry)
XX Rat brain homologue of serine ectopeptidase.
XX Serine ectopeptidase; tripeptidylpeptidase II; TPP II;
XX alternative splicing; rodent; rat; cerebral cortex;
XX neuropeptide; cholecystokinin; CCK; inactivation; degradation;
XX anorexia; schizophrenia; Parkinson's disease; depression;
XX irritable bowel syndrome; bulimia; pathological obesity; ss.
XX Rattus sp.
XX OS
XX Key 599..4306
XX CDS /*tag= a
XX /*product= serine_ectopeptidase
XX 849..850
XX misc_recomb /*tag= b
XX /*label= Alternative_splice_site
XX /*note= "A second clone was isolated which differed
XX upstream of this site and was identical
XX downstream of it (see T72653)."
XX
XX W09635805-A2.
XX 14-NOV-1996.
XX 09-MAY-1996; 96WO-FR00700.
XX 09-MAY-1995; 95FR-0005489.
XX (INRM ) INSERM-INST SANTE & RECH MEDICALE.
XX Bamba RB, Bishop PB, Bourgeat P, Chan S, Ganellin CR;
XX Leblond B, Moire ANJ, Schwartz JC, Vargas F;
XX Lihua Z, Rose C;
XX WPI; 1996-518693/51.
XX P-PSDB; W21572.
XX Screening medicaments for treating disorders linked to inactivation
XX of endogenous neuro-peptide(s) - by contacting candidate molecule
XX with membrane tri-peptidyl-peptidase (homologue) and measuring
XX enzyme activity
XX Disclosure; Pages 166-171; 212pp; French.
XX An enzyme with specificity for cholecystokinin (CCK) substrates
XX CC
```

CC (specifically the non-sulphated CCK8 and the CCK5 peptides) was  
 CC purified from rat cerebral cortex membranes using high performance  
 CC liquid chromatography. Based on amino acid sequence data from the  
 CC purified enzyme, PCR primers were designed to amplify oligonucleotide  
 CC probes A and B of lengths 350 and 380 nucleotides, respectively.  
 CC Using these probes, two distinct clones were identified in a rat  
 CC brain cDNA library. The first clone had the sequence given in  
 CC T72053 and is the rodent homologue of human tripeptidylpeptidase II.  
 CC The second clone had the present sequence which differs from the  
 CC first clone in the 5' region, probably as a result of alternative  
 CC splicing, and appears to encode a serine ectopeptidase. The rat enzymes  
 CC are preferred for use in a new method of screening for medicaments  
 CC for treating disorders linked to the inactivation or degradation (or  
 CC being treatable by retarding physiological degradation) of endogenous  
 CC neuropeptides. In the method, a candidate molecule is contacted with  
 CC a membrane tripeptidylpeptidase or homologue, and enzyme activity is  
 CC measured. Disorders linked to inactivation or degradation of endogenous  
 CC neuropeptides, include food intake disorders, cognitive and motor  
 CC disorders such as anorexia, schizophrenia, Parkinson's disease and  
 CC depression, as well as gastrointestinal transit disorders such as  
 CC irritable bowel syndrome, bulimia and pathological obesity.  
 XX  
 SQ Sequence 5109 BP; 1541 A; 969 C; 1061 G; 1538 T; 0 other;

Query Match 3.3%; Score 33; DB 17; Length 5109;  
 Best Local Similarity 100.0%; Pred. NO. 1e-06;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 284 ccattacagatggttgagccaccatggtt 316  
 Db 416 ccattacagatggttgagccaccatggtt 448

RESULT 9

O05243/c  
 ID Q05243 standard; DNA; 13206 BP.

AC Q05243;

DT 29-NOV-1990 (first entry)

DE Rat acyl peptide hydrolase (APH) gene and flanking regions.

KW Acyl peptide hydrolase; APH; cancer; small cell carcinoma; SCC.

XX Synthetic.

Key	Location/Qualifiers
FT exon	2580..2842
FT intron	/*tag= a
FT exon	2843..3072
FT intron	/*tag= b
FT exon	3073..3216
FT intron	/*tag= c
FT exon	3217..3290
FT intron	/*tag= d
FT exon	3291..3423
FT intron	/*tag= e
FT exon	3424..3875
FT intron	/*tag= f
FT exon	3876..4002
FT intron	/*tag= g
FT exon	4003..4748
FT intron	/*tag= h
FT exon	4749..4842
FT intron	/*tag= i
FT exon	4843..4934
FT intron	/*tag= j
FT exon	4935..5010
FT intron	/*tag= k
FT exon	5011..5099
FT intron	/*tag= l

FT	exon	FT
5100..5263	/*tag= m	exon
5264..5404	/*tag= n	intron
5405..5542	/*tag= o	exon
5542..5636	/*tag= p	intron
5637..5728	/*tag= q	exon
5729..5804	/*tag= r	intron
5805..5845	/*tag= s	exon
5846..5936	/*tag= t	intron
5937..6058	/*tag= u	exon
6059..6520	/*tag= v	intron
6521..6581	/*tag= w	exon
6582..7584	/*tag= x	intron
7585..7682	/*tag= y	exon
7683..8283	/*tag= z	intron
8284..8335	/*tag= aa	exon
8336..9439	/*tag= ab	intron
9440..9528	/*tag= ac	exon
9529..9886	/*tag= ad	intron
9887..10025	/*tag= ae	exon
10026..10436	/*tag= af	intron
10437..10520	/*tag= ag	exon
10521..10668	/*tag= ah	intron
10669..10749	/*tag= ai	exon
10750..11203	/*tag= aj	intron
11204..11292	/*tag= ak	exon
11293..11385	/*tag= al	intron
11386..11576	/*tag= am	exon
11577..11660	/*tag= an	intron
11661..11763	/*tag= ao	exon
11764..11850	/*tag= ap	intron
11851..11957	/*tag= aq	exon
11958..12042	/*tag= ar	intron
12043..12304	/*tag= as	exon
1664..1859	/*tag= at	repeat_unit
1860..2047	/*tag= au	repeat_unit

EP378224-A.

XX



us-09-101-423a-3.olig.rng

Mon Nov 6 10:14:17 2000

CC The present sequence represents cosmid CVO14, which contains the rat  
 CC vasopressin and oxytocin locus. The specification describes 5'-OT-EST  
 CC (oxytocin expressed sequence tag) polypeptides. The 5'-OT EST gene is  
 CC involved in the control of obesity and fertility in males. 5'-OT EST  
 CC nucleic acids are useful for producing transgenic animals. The  
 CC transgenic animals created serve as a model for human late onset  
 CC obesity and other related disorders and are also used for identifying  
 CC the genetic cause of obesity. Compounds which modulate 5'-OT EST  
 CC expression or activity are useful in the treatment or modulation of  
 CC late onset visceral obesity or male infertility particularly in the  
 CC disorders related to these conditions such as wasting, or anorexia,  
 CC or cachexia associated with prolonged illness, or malabsorptive states  
 CC or catabolic states associated with other diseases such as inflammatory  
 CC conditions, Crohn's disease or AIDS wasting, or burns, or cancer, or  
 CC bone disease.  
 CC  
 CC Sequence 44576 BP; 12157 A; 10993 C; 10857 G; 10569 T; 0 other;  
 XX  
 SQ

Query Match 3.2%; Score 32; DB 21; Length 44576;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-06;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 cattacagatggttgagccaccatgtggtt 316  
 ||||||||||||||||||||||||||||||||  
 DB 31688 cattacagatggttgagccaccatgtggtt 31719

RESULT 11  
 Z23891  
 ID Z23891 standard; DNA; 49999 BP.  
 XX  
 AC Z23891;  
 XX  
 DT 25-JAN-2000 (first entry)  
 XX  
 DE Murine LOBO genomic DNA fragment 1.  
 XX  
 KW LOBO; long bones; bone development; bone extension; skull; osteopathic;  
 KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
 KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO9950284-A2.  
 XX  
 PD 07-OCT-1999.  
 XX  
 PF 26-MAR-1999; 99WO-EP02055.  
 XX  
 PR 27-MAR-1998; 98DE-1013799.  
 XX  
 PA (ROSE/) ROSENTHAL A.  
 XX  
 PI Rosenthal A, Fump A, Hess J, Aigner T, Wirth T;  
 XX  
 XX WPI; 1999-601320/51.  
 DR  
 XX  
 XX Nucleic acids encoding proteins which influence bone development,  
 PT useful for treating and studying bone disorders -  
 XX  
 PS Example 3; Page 69-97; 391pp; German.  
 XX  
 CC This invention describes novel nucleic acids (I; designated LOBO (long  
 CC bones)) encoding proteins influencing bone development in mammals. The  
 CC proteins of the invention reduce and/or inactivate bone extension (i.e.  
 CC development), with exception of the skull and have osteopathic activity.  
 CC The nucleic acid molecules, proteins and antibodies can be used in  
 CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
 CC and nucleic acid molecules, etc. are useful for production of transgenic  
 CC animals, especially a transgenic mouse for the study of diseases  
 CC associated with bone development, e.g. spondyloepiphyseal dysplasia and  
 CC achondroplasia. This sequence encodes the murine LOBO protein described

PD 18-JUL-1990.  
 XX  
 PF 12-JAN-1990; 90EP-0100575.  
 XX  
 PR 13-JAN-1989; 89US-0296996.  
 PR 01-NOV-1989; 89US-0429935.  
 XX  
 PA (GEO-) GEN HOSPITAL CORP.  
 XX  
 XX Smith JA;  
 PI  
 XX WPI; 1990-218599/29.  
 DR  
 XX Detecting small cell carcinoma by hybridisation and analysis -  
 PT using nucleic acid encoding acyl-peptide hydrolase enzyme.  
 XX  
 PS Disclosure; 47pp; English.  
 CC  
 CC APH may be used in hydrolysis of an N-alpha-acetylated peptide,  
 CC Probes raised to the peptide product may be used to detect and  
 CC identify SCC and other cancers arising from chromosome 3 deletions.  
 CC See also Q05241.  
 XX  
 SQ Sequence 13206 BP; 3039 A; 3419 C; 3494 G; 3254 T; 0 other;  
 XX

Query Match 3.2%; Score 32; DB 11; Length 13206;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-06;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 cattacagatggttgagccaccatgtggtt 316  
 ||||||||||||||||||||||||||||||||  
 DB 4528 cattacagatggttgagccaccatgtggtt 4497

RESULT 10  
 Z61522  
 ID Z61522 standard; DNA; 44576 BP.  
 XX  
 AC Z61522;  
 XX  
 DT 19-JUN-2000 (first entry)  
 XX  
 DE Cosmid CVO14 containing rat vasopressin and oxytocin locus.  
 XX  
 KW Oxytocin expressed sequence tag; 5'-OT EST; obesity; fertility; male;  
 KW transgenic animal; human late onset obesity; late onset visceral obesity;  
 KW male infertility; wasting; anorexia; cachexia; malabsorptive state;  
 KW catabolic state; inflammatory condition; Crohn's disease; AIDS wasting;  
 KW burn; cancer; bone disease; vasopressin; oxytocin; ss.  
 XX  
 OS Rattus sp.  
 XX  
 PN WO200009686-A1.  
 XX  
 PD 24-FEB-2000.  
 XX  
 PF 12-AUG-1999; 99WO-GB02658.  
 XX  
 PR 12-AUG-1998; 98GB-0017566.  
 PR 06-MAY-1999; 99GB-0010522.  
 XX  
 PA (MEDI-) MEDICAL RES COUNCIL.  
 XX  
 XX Robinson ICAF, Stoye JP, Flavell D, Wells SE, Le Tissier P;  
 PI WPI; 2000-224331/19.  
 DR  
 XX New anti-obesity polypeptide useful for treating obesity or infertility  
 PT in mammals -  
 XX  
 PS Claim 9; Page 129-154; 162pp; English.  
 XX

CC in the method of the invention.

XX Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;

SQ Query Match 3.2%; Score 32; DB 20; Length 49999;

Best Local Similarity 100.0%; Pred. No. 3.1e-06;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 285 cattacagatggttgagccaccatgtggtt 316

Db 30947 cattacagatggttgagccaccatgtggtt 30978

RESULT 12

223896

ID 223896 standard; DNA; 49999 BP.

XX AC 223896;

XX 223896;

DT 25-JAN-2000 (first entry)

XX Murine LOBO homologue genomic DNA fragment 2.

DE LOBO; long bones; bone development; bone extension; skull; osteopathic;  
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.

XX Mus musculus.

OS WO9950284-A2.

XX 07-OCT-1999.

XX 26-MAR-1999; 99WO-EP02055.

XX 27-MAR-1998; 98DE-1013799.

XX (ROSE/) ROSENTHAL A.

XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;

XX WPI; 1999-601320/51.

XX Nucleic acids encoding proteins which influence bone development,

XX useful for treating and studying bone disorders -

XX Example 3; Page 161-189; 391pp; German.

XX This invention describes novel nucleic acids (I; designated LOBO (long

XX bones)) encoding proteins influencing bone development in mammals. The

XX proteins of the invention reduce and/or inactivate bone extension (i.e.

XX development), with exception of the skull and have osteopathic activity.

XX The nucleic acid molecules, proteins and antibodies can be used in

XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods

XX and nucleic acid molecules, etc. are useful for production of transgenic

XX animals, especially a transgenic mouse for the study of diseases

XX associated with bone development, e.g. spondyloepiphyseal dysplasia and

XX achondroplasia. This sequence encodes the murine LOBO protein described

XX in the method of the invention.

XX Sequence 49999 BP; 13135 A; 11787 C; 10868 G; 14209 T; 0 other;

SQ Query Match 3.2%; Score 32; DB 20; Length 49999;

Best Local Similarity 100.0%; Pred. No. 3.1e-06;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 285 cattacagatggttgagccaccatgtggtt 316

Db 32930 cattacagatggttgagccaccatgtggtt 32961

RESULT 13

X83007/C

ID X83007 standard; DNA; 51259 BP.

XX AC X83007;

XX 31-AUG-1999 (first entry)

DT Partial mouse WRN genomic sequence #3.

DE Mouse; WRN; Werner's syndrome; detection; diagnosis; autosomal;

XX recessive disorder; phenotype; ss.

KW Mus musculus.

OS WO9724435-A1.

XX 10-JUL-1997.

XX 30-DEC-1996; 96WO-US20785.

XX 12-APR-1996; 96US-0632175.

XX 29-DEC-1995; 95US-0009409.

XX 29-DEC-1995; 95US-0800539.

XX 30-JAN-1996; 96US-0010835.

XX 30-JAN-1996; 96US-0594242.

XX (DARW-) DARWIN MOLECULAR CORP.

XX (OSHI/) OSHIMA J.

XX Fu Y, Mulligan J, Oshima J, Schellenberg GD, Yu C;

XX WPI; 1997-363671/33.

XX Isolated nucleic acid molecule encoding the WRN gene product -

XX useful for detection and treatment of Werner's syndrome, and related

XX diseases

XX Claim 1; Fig 7; 153pp; English.

XX This sequence represents a fragment of the genomic sequence containing

XX the coding region for the mouse WRN gene (X83004). The corresponding

XX human gene (X83001) encodes a protein related to Werner's syndrome.

XX The products can be used for the detection and treatment of Werner's

XX syndrome (WS), an autosomal recessive disorder with a complex phenotype,

XX as well as related diseases.

XX Sequence 51259 BP; 14533 A; 9635 C; 10266 G; 16825 T; 0 other;

SQ Query Match 3.2%; Score 32; DB 18; Length 51259;

Best Local Similarity 100.0%; Pred. No. 3.1e-06;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 285 cattacagatggttgagccaccatgtggtt 316

Db 6380 CATTACAGATGTTGTGAGCCACCATGTGTT 6349

RESULT 14

T08126

ID T08126 standard; DNA; 26698 BP.

XX AC T08126;

XX 13-AUG-1996 (first entry)

DT Mouse syndecan-1 gene.

DE Syndecan-1; ectodomain; tumour; cancer; therapy; proteoglycan;

XX cell differentiation; ss.

KW Mus sp.

XX OS

ID	XX	Q67902	standard; DNA; 26700 BP.
XX	AC	Q67902;	
XX	DT	08-DEC-1994 (first entry)	
XX	DE	Syndecan gene.	
XX	DE	Syndecan; enhancer; differentiation; tumor; therapy; hair growth; ss.	
XX	KW	Mus sp.	
XX	OS	Location/Qualifiers	
XX	XX	Key	
XX	FT	intron	1..4377
XX	FT	exon	/*tag= a
XX	FT	exon	4378..4443
XX	FT	intron	/*tag= b
XX	FT	intron	4444..22025
XX	FT	exon	/*tag= c
XX	FT	exon	22026..22106
XX	FT	intron	/*tag= d
XX	FT	intron	22107..23000
XX	FT	exon	/*tag= e
XX	FT	exon	23001..23483
XX	FT	intron	/*tag= f
XX	FT	intron	23484..23904
XX	FT	exon	/*tag= g
XX	FT	exon	23905..24039
XX	FT	intron	/*tag= h
XX	FT	intron	24040..24250
XX	FT	exon	/*tag= i
XX	FT	exon	24251..24418
XX	FT	intron	/*tag= j
XX	FT	intron	24419..26700
XX	FT	intron	/*tag= k
XX	PN	WO9412162-A.	
XX	XX	09-JUN-1994.	
XX	PD	01-DEC-1993; 93WO-FI00514.	
XX	XX	01-DEC-1992; 92US-0988427.	
XX	XX	(WAER/) WAERRI A.M.	
XX	PA	(ALAN/) ALANEN-KURKI L M.	
XX	PA	(AUVI/) AUVINEN P O V.	
XX	PI	Alanen-Kurki LM; Auvinen POV, Jaakkola PM, Jalkanen MT;	
XX	PI	Leppaesm, Mali MS, Vihinen TA, Waerri AM;	
XX	XX	WPI: 1994-199926/24.	
XX	DR	P-PSDB; R55276.	
XX	XX	Syndecan stimulation of cellular differentiation - useful for	
XX	PT	decreasing tumour growth used to promote hair growth	
XX	PS	Disclosure; Page 22-39; 65pp; English.	
XX	XX	The mouse syndecan gene enhancer, located 8-10 kb upstream from the	
XX	CC	initiation site; is given in Q67901. Manipulation of the enhancer	
XX	CC	can be used either to slow or prevent tumor growth or to promote	
XX	CC	differentiation of specific cell types, e.g. epidermal cells to	
XX	CC	promote hair formation. The complete mouse syndecan gene and its	
XX	CC	encoded protein are given in Q67902 and R55276.	
XX	XX	Sequence 26700 BP; 5742 A; 6559 C; 7233 G; 7165 T; 1 other;	

```

Query Match      3.1%; Score 31; DB 15; Length 26700;
Best Local Similarity 100.0%; Pred. No. 9.8e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Mon Nov 6 10:14:17 2000

us-09-101-423a-3.olig.rng

Page 10

Qy 364 gaccttggagagcagtcagtgccttaac 394  
|||||  
Db 1627 gaccttggagagcagtcagtgccttaac 1657

Search completed: November 4, 2000, 03:38:12  
Job time: 11156 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 4, 2000, 03:31:32 ; Search time 140.17 Seconds  
(without alignments)  
1087.618 Million cell updates/sec

Title: US-09-101-423A-3  
Perfect score: 1008  
Sequence: 1 agggaccaggttcacatccc.....ttctgtgtctctctggaac 1008

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 262060 seqs, 75620727 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/5E\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/5F\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/5G\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	3.2	51259	5	US-08-781-891-209
2	31	3.1	26700	2	US-08-472-217-1
3	31	3.1	26700	3	US-08-488-199-5
4	31	3.1	26700	4	US-08-760-334A-1
5	30	3.0	13011	4	US-08-791-849A-14
6	27	2.7	6727	5	US-08-629-643A-5
7	27	2.7	13146	4	US-08-724-354D-3
8	27	2.7	13146	5	US-09-270-984A-3
9	26	2.6	2751	2	US-08-153-848-23
10	26	2.6	2751	5	US-09-239-843A-23
11	26	2.6	2751	6	PCT-US93-11153-23
12	26	2.6	7970	1	US-08-135-511-31
13	26	2.6	7997	2	US-08-187-453-31
14	26	2.6	7997	2	US-08-562-985A-7
15	25	2.5	601	5	US-09-080-855-34
16	25	2.5	613	5	US-09-080-855-35
17	25	2.5	2294	3	US-08-676-279-49
18	22	2.2	4765	5	US-08-936-135-21
19	22	2.2	4780	5	US-08-936-135-23
20	21	2.1	2498	1	US-07-943-843-3
21	21	2.1	2498	2	US-08-347-003-3
22	20	2.0	11236	1	US-07-853-913-1
23	19	1.9	2720	3	US-08-007-107-3
24	19	1.9	4698	1	US-07-807-043B-5
25	19	1.9	4698	1	US-08-299-849B-5
26	19	1.9	4698	4	US-08-142-368A-5

Sequence 5, Appl1  
Sequence 1, Appl1  
Sequence 1, Appl1  
Sequence 209, Appl1  
Sequence 3, Appl1  
Sequence 2, Appl1  
Sequence 5, Appl1  
Patent No. 5258287  
Sequence 4, Appl1  
Sequence 4, Appl1  
Sequence 1, Appl1  
Sequence 3, Appl1  
Sequence 3, Appl1  
Sequence 1, Appl1  
Sequence 1, Appl1  
Sequence 1, Appl1  
Patent No. 5453251  
Sequence 5, Appl1  
Sequence 5, Appl1

ALIGNMENTS

RESULT 1  
US-08-781-891-209/c  
; Sequence 209, Application US/08781891  
; Patent No. 609062c  
; GENERAL INFORMATION:  
; APPLICANT: Fu, Ying-Hui  
; APPLICANT: Yu, Chang-En  
; APPLICANT: Ostima, Junko  
; APPLICANT: Mulligan, John T.  
; APPLICANT: Schellenberg, Gerald D.  
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
; TITLE OF INVENTION: WERNER'S SYNDROME  
; NUMBER OF SEQUENCES: 209  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 630C Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IEM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 209:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51259 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-209

Query Match 3.2%; Score 32; DB 5; Length 51259;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 285 cattacagatgtgtgagccaccatgtggtt 316  
|||||  
Db 6380 CATTACAGATGTTGTGAGCCACCAATGTGTT 6349

RESULT 2  
US-08-472-217-1  
; Sequence 1, Application US/08472217  
; Patent No. 5726058  
; GENERAL INFORMATION:  
; APPLICANT: Alanen-Kurki, Leena  
; APPLICANT: Auvinen, Petri  
; APPLICANT: Jaakkola, Panu  
; APPLICANT: Jalkanen, Markku  
; APPLICANT: Lepp, Sirpa  
; APPLICANT: Mali, Markku  
; APPLICANT: Vihinen, Tapani  
; APPLICANT: W rrl, Anni  
; TITLE OF INVENTION: Syndecan Stimulation Of Cellular  
; TITLE OF INVENTION: Differentiation  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/472.217  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/206,186  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/988,427  
; FILING DATE: 01-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cimbal, Michele A.  
; REGISTRATION NUMBER: 33,851  
; REFERENCE/DOCKET NUMBER: 1102.0050003  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26700 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: Join(4378..4443, 22026..22106, 23001..23483,  
; LOCATION: 23905..24039, 24251..24418)  
US-08-472-217-1

Query Match 3.1%; Score 31; DB 2; Length 26700;  
Best Local Similarity 100.0%; Pred. No. 3.8e-06;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 364 gacatttggaagagcagtcagtccttaac 394  
|||||  
Db 1627 GACCTTGGAGAGCAGTCAGTCCTTAAC 1657

RESULT 3  
US-08-488-199-5  
; Sequence 5, Application US/08488199  
; Patent No. 5851993  
; GENERAL INFORMATION:  
; APPLICANT: Jalkanen, Markku  
; APPLICANT: Mali, Markku  
; TITLE OF INVENTION: Suppression of Tumor Cell Growth By  
; TITLE OF INVENTION: Syndecan-1 Ectodomain  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Ave., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,199  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/258,862  
; FILING DATE: 13-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cimbal, Michele A.  
; REGISTRATION NUMBER: 33,851  
; REFERENCE/DOCKET NUMBER: 1102.0130001  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26700 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 4378..4443  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 22026..22107  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 23002..23483  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 23905..24040  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 24252..24418  
US-08-488-199-5

Query Match 3.1%; Score 31; DB 3; Length 26700;  
Best Local Similarity 100.0%; Pred. No. 3.8e-06;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 364 gacatttggaagagcagtcagtccttaac 394  
|||||  
Db 1627 GACCTTGGAGAGCAGTCAGTCCTTAAC 1657

RESULT 4  
US-08-760-534A-1

us-09-101-423a-3.olig.rni

Mon Nov 6 10:14:18 2000

```

; Sequence 1, Application US/08760534A
; Patent No. 6017727
; GENERAL INFORMATION:
; APPLICANT: JAAKANEN, MARKKU
; APPLICANT: JAAKKOLA, PANU
; APPLICANT: VIHINEN, TAPANI
; TITLE OF INVENTION: SYNDCAN ENHANCER ELEMENT AND SYNDECAN
; TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/760,534A
; FILING DATE: 02-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/206,186
; FILING DATE: 07-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/PI93/00514
; FILING DATE: 01-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: CIMBALA, MICHELE A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1708.0050004/MAC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4378..4443, 22026..22106, 23001..23483,
; 23905..24039, 24251..24418)
; US-08-760-534A-1

Query Match 3.1%; Score 31; DB 5; Length 26700;
Best Local Similarity 100.0%; Pred. No. 3.8e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 gaccttggagagcagtcagtccttaac 394
Db 1627 GACCTTTGGAAGAGCAGTCAGTCCTTAAC 1657

RESULT 5
US-08-791-849A-14
; Sequence 14, Application US/08791849A
; Patent No. 5914449
; GENERAL INFORMATION:
; APPLICANT: Makoto MURASE et al.
; TITLE OF INVENTION: Method for Increasing Storage
; TITLE OF INVENTION: Lipid Content in Plant Seed
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700

```

```

; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791,849A
; FILING DATE: January 30, 1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13011 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: rat (Rattus norvegicus)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(3212..3218, 3766..3948, 5917..6008,
; 6152..6283, 6418..6604, 6921..7191, 7302..7452, 7665..7817, 7911..8
; 9298..9479, 10163..10269)
; FEATURE:
; NAME/KEY: exon
; LOCATION: join(3194..3218, 3766..3948, 5917..6008,
; 6152..6283, 6418..6604, 6921..7191, 7302..7452, 7665..7817, 7911..8
; 9480..10162)
; NAME/KEY: intron
; LOCATION: join(3219..3765, 3949..5916, 6009..6151,
; 6284..6417, 6605..6920, 7191..7301, 7453..7664, 7818..7910, 8078..9
; 9480..10162)
; US-08-791-849A-14

Query Match 3.0%; Score 30; DB 4; Length 13011;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ttacagatggttgagccaccatgtggtt 316
Db 9711 TTACAGATGTTGTGAGCCACCATGTGTT 9740

RESULT 6
US-08-629-643A-5
; Sequence 5, Application US/08629643A
; Patent No. 6025539
; GENERAL INFORMATION:
; APPLICANT: Lee, J. L.
; APPLICANT: Lee, N. A.
; TITLE OF INVENTION: IL-5 TRANSGENIC MOUSE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.
; STREET: P. C. Box 2938
; CITY: Minneapolis

```

STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/629,643A  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/629,643  
FILING DATE: 09-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Viksnins, Ann S  
REGISTRATION NUMBER: 37,748  
REFERENCE/DOCKET NUMBER: 150.167W01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-359-3260  
TELEFAX: 612-359-3263  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6727 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 650..3771  
OTHER INFORMATION: Join 650..730, 1560..1592,  
OTHER INFORMATION: 3468..3596, 3676..3771  
US-08-629-643A-5

Query Match 2.7%; Score 27; DB 5; Length 6727;  
Best Local Similarity 100.0%; Pred. No. 0.00035;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 ggaattgaactcaggacaccttgggaaga 376  
|||||  
Db 6196 GGAATTGACTCAGGACCTTTGGAAGA 6222

RESULT 7  
US-08-724-354D-3  
Sequence 3, Application US/08724354D  
Patent No. 5994119  
GENERAL INFORMATION:  
APPLICANT: Dietz, Harry C.  
TITLE OF INVENTION: MAMMALIAN REGULATOR OF  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,354D  
FILING DATE: 01-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/016,482

FILING DATE: 29-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/090001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13146 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-08-724-354D-3

Query Match 2.7%; Score 27; DB 4; Length 13146;  
Best Local Similarity 100.0%; Pred. No. 0.00035;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 cagatggttgagaccaccatgtggtt 316  
|||||  
Db 819 CAGATGTTGTGAGCCACCACATGTGTT 845

RESULT 8  
US-09-270-984A-3  
Sequence 3, Application US/09270984A  
Patent No. 6048965  
GENERAL INFORMATION:  
APPLICANT: Dietz, Harry C.  
TITLE OF INVENTION: MAMMALIAN REGULATOR OF  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/270,984A  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/724,354  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07265/090001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13146 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-09-270-984A-3

Query Match 2.7%; Score 27; DB 5; Length 13146;  
Best Local Similarity 100.0%; Pred. No. 0.00035;



us-09-101-423a-3.olig.rni

Mon Nov 6 10:14:18 2000

MATCHES 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 cagatggttgagccacacatgtggtt 316  
 DB 819 CAGATGTTGTGAGCCACCATGTGTT 845

RESULT 9  
 US-08-153-848-23  
 ; Sequence 23, Application US/08153848  
 ; Patent No. 5759804  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godiska, Ronald  
 ; APPLICANT: Gray, Patrick W.  
 ; APPLICANT: Schwelkart, Vicki L.  
 ; TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors  
 ; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
 ; ADDRESSEE: Ricknell  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/153,848  
 ; FILING DATE:  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/977,452  
 ; FILING DATE: 17-NOV-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: No. 5759804and, Greta E.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 31794  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (312) 474-6300  
 ; TELEFAX: (312) 474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2751 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 1..691  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 692..1771  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 692..1768  
 ; FEATURE:  
 ; NAME/KEY: polyA\_signal  
 ; LOCATION: 2341..2348  
 ; US-08-153-848-23

Query Match 2.6%; Score 26; DB 2; Length 2751;  
 Best Local Similarity 100.0%; Pred. No. 0.0011;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 cccattacagatggttgagccacc 308  
 |||||||||||||||||||||||||||

DB 187 CCCATTACAGATGTTGTGAGCCACC 212

RESULT 10  
 US-09-299-843A-23  
 ; Sequence 23, Application US/09299843A  
 ; Patent No. 6107475  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godiska, Ronald  
 ; APPLICANT: Gray, Patrick W.  
 ; APPLICANT: Schwelkart, Vicki L.  
 ; TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors  
 ; NUMBER OF SEQUENCES: 66  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
 ; ADDRESSEE: Eorun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/299,843A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 09/088,337  
 ; FILING DATE: 01-JUN-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/153,848  
 ; FILING DATE: 17-NOV-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/977,452  
 ; FILING DATE: 17-NOV-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jill F. Uhl  
 ; REGISTRATION NUMBER: 43,213  
 ; REFERENCE/DOCKET NUMBER: 27866/32059B  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (312) 474-6300  
 ; TELEFAX: (312) 474-0448  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2751 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 1..691  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 692..1771  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 692..1768  
 ; FEATURE:  
 ; NAME/KEY: polyA\_signal  
 ; LOCATION: 2341..2348  
 ; US-09-299-843A-23

Query Match 2.6%; Score 26; DB 5; Length 2751;  
 Best Local Similarity 100.0%; Pred. No. 0.0011;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 ccattacagatgtgtgagccacc 308  
 |||  
 Db 187 CCATTACAGATGTTGTGAGCCACC 212

RESULT 11  
 PCT-US93-11153-23  
 ; Sequence 23, Application PC/TUS9311153  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godiska, Ronald  
 ; APPLICANT: Gray, Patrick W.  
 ; APPLICANT: Schweikert, Vicki L.  
 ; TITLE OF INVENTION: Novel Seven Transmembrane Receptors  
 ; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
 ; ADDRESSEE: Bicknell  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/11153  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/977,452  
 ; FILING DATE: 17-NOV-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Noland, Greta E.  
 ; REGISTRATION NUMBER: 35,302  
 ; REFERENCE/DOCKET NUMBER: 31794  
 ; TELEPHONE: (312) 474-6300  
 ; TELEFAX: (312) 474-0448  
 ; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2751 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 1..691  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 692..1771  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 692..1768  
 ; FEATURE:  
 ; NAME/KEY: polyA\_signal  
 ; LOCATION: 2341..2348

PCT-US93-11153-23  
 Query Match 2.6%; Score 26; DB 6; Length 2751;  
 Best Local Similarity 100.0%; Pred. No. 0.0011;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 283 ccattacagatgtgtgagccacc 308  
 |||  
 Db 187 CCATTACAGATGTTGTGAGCCACC 212

RESULT 12  
 US-08-135-511-31  
 ; Sequence 31, Application US/08135511  
 ; Patent No. 5558999  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chiang, John  
 ; TITLE OF INVENTION: Cholesterol 7a-Hydroxylase Gene  
 ; TITLE OF INVENTION: Regulatory Elements and Methods for Using Them  
 ; NUMBER OF SEQUENCES: 35  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington, D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/135,511  
 ; FILING DATE: 13-OCT-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: SANDEROCK, Colin G.  
 ; REGISTRATION NUMBER: 31,298  
 ; REFERENCE/DOCKET NUMBER: 18748/175  
 ; TELEPHONE: (302) 672-5300  
 ; TELEFAX: (302) 672-5399  
 ; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 31:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 7970 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-135-511-31

Query Match 2.6%; Score 26; DB 1; Length 7970;  
 Best Local Similarity 100.0%; Pred. No. 0.0011;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 287 ttacagatgtgtgagccaccatgt 312  
 |||  
 Db 5316 TTACAGATGTTGTGAGCCACCATGT 5341

RESULT 13  
 US-08-187-453-31  
 ; Sequence 31, Application US/08187453  
 ; Patent No. 5753431  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chiang, John  
 ; TITLE OF INVENTION: Cholesterol 7a-Hydroxylase Gene  
 ; TITLE OF INVENTION: Regulatory Elements and Transcription Factors  
 ; NUMBER OF SEQUENCES: 37  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington, D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/187,453

us-09-101-423a-3.olig.rni

Mon Nov 6 10:14:18 2000

```
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/135,488
; FILING DATE: 13-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/135,511
; FILING DATE: 13-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/135,510
; FILING DATE: 13-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SANDERCOCK, Colin G.
; REGISTRATION NUMBER: 31,298
; REFERENCE/DOCKET NUMBER: 18748/188
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7997 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-187-453-31

Query Match 2.6%; Score 26; DB 2; Length 7997;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 ttacagatggttgagccaccatgt 312
|||||
DB 5343 TTACAGATGTTGTGAGCCACCACGT 5368

RESULT 15
US-09-080-855-34
; Sequence 34, Application US/09080855A
; Patent No. 6083721
; GENERAL INFORMATION:
; APPLICANT: Saras, Jan
; APPLICANT: Franz, Petra
; APPLICANT: Aspenstrm, Pontus
; APPLICANT: Hellman, Ulf
; APPLICANT: Gomez, Leonel Jorge
; APPLICANT: Heidt, Carl-Henrik
; TITLE OF INVENTION: PARG, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL1
; FILE REFERENCE: I0461/7030
; CURRENT APPLICATION NUMBER: US/09/080,855A
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: 08/805,583
; EARLIER FILING DATE: 1997-02-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 493..493
; OTHER INFORMATION: n = a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 535..535
; OTHER INFORMATION: n = a, c, g or t
US-09-080-855-34

Query Match 2.5%; Score 25; DB 5; Length 601;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 gatggttgtagccaccatgtggtt 316
|||||
DB 59 gatggttgtagccaccatgtggtt 83

Search completed: November 4, 2000, 03:32:45
Job time: 11955 sec
```

Mon Nov 6 10:14:18 2000

us-09-101-423a-3.olig.rni

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 01:08:41 ; Search time 3299.68 Seconds  
(without alignments)  
1888.749 Million cell updates/sec

Title: US-09-101-423A-3  
Perfect score: 1008  
Sequence: 1 agaccagagttccatcc.....ttcctgtgtccctcctggaac 1008

Scoring table: OLIGO\_NVC  
Gapop 60.0 , Gapext 60.0

Searched: 7189864 seqs, 3091403243 residues

Word size : 0

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*
- 10: gb\_est10.\*
- 11: gb\_est11.\*
- 12: gb\_est12.\*
- 13: gb\_est13.\*
- 14: gb\_est14.\*
- 15: gb\_est15.\*
- 16: gb\_est16.\*
- 17: gb\_est17.\*
- 18: gb\_est18.\*
- 19: gb\_est19.\*
- 20: gb\_est20.\*
- 21: gb\_est21.\*
- 22: gb\_est22.\*
- 23: gb\_est23.\*
- 24: gb\_est24.\*
- 25: gb\_est25.\*
- 26: gb\_est26.\*
- 27: gb\_est27.\*
- 28: gb\_est28.\*
- 29: gb\_est29.\*
- 30: gb\_est30.\*
- 31: gb\_est31.\*
- 32: gb\_est32.\*
- 33: gb\_est33.\*
- 34: gb\_est34.\*
- 35: gb\_est35.\*
- 36: gb\_est36.\*
- 37: gb\_est37.\*
- 38: gb\_est38.\*
- 39: gb\_est39.\*
- 40: gb\_est40.\*
- 41: em\_estba.\*
- 42: em\_estfun.\*
- 43: em\_esthum1.\*

- 44: em\_esthum2.\*
- 45: em\_esthum3.\*
- 46: em\_esthum4.\*
- 47: em\_esthum5.\*
- 48: em\_esthum6.\*
- 49: em\_esthum7.\*
- 50: em\_esthum8.\*
- 51: em\_esthum9.\*
- 52: em\_esthum10.\*
- 53: em\_esthum11.\*
- 54: em\_esthum12.\*
- 55: em\_esthum13.\*
- 56: em\_esthum14.\*
- 57: em\_esthum15.\*
- 58: em\_esthum16.\*
- 59: em\_esthum17.\*
- 60: em\_esthum18.\*
- 61: em\_esthum19.\*
- 62: em\_esthum20.\*
- 63: em\_estin1.\*
- 64: em\_estin2.\*
- 65: em\_estin3.\*
- 66: em\_estin4.\*
- 67: em\_estom.\*
- 68: em\_estov1.\*
- 69: em\_estov2.\*
- 70: em\_estpl1.\*
- 71: em\_estpl2.\*
- 72: em\_estpl3.\*
- 73: em\_estpl4.\*
- 74: em\_estpl5.\*
- 75: em\_estrol.\*
- 76: em\_estro2.\*
- 77: em\_estro3.\*
- 78: em\_estro4.\*
- 79: em\_estro5.\*
- 80: em\_estro6.\*
- 81: em\_estro7.\*
- 82: em\_estro8.\*
- 83: em\_estro9.\*
- 84: em\_estro10.\*
- 85: em\_estro11.\*
- 86: em\_estro12.\*
- 87: em\_estro13.\*
- 88: gb\_gss1.\*
- 89: gb\_gss2.\*
- 90: gb\_gss3.\*
- 91: gb\_gss4.\*
- 92: em\_gss1.\*
- 93: em\_gss2.\*
- 94: em\_gss3.\*
- 95: em\_gss4.\*
- 96: gb\_gss5.\*
- 97: gb\_gss6.\*
- 98: gb\_gss7.\*
- 99: gb\_gss8.\*
- 100: gb\_gss9.\*
- 101: em\_gss5.\*
- 102: em\_gss6.\*
- 103: em\_gss7.\*
- 104: em\_gss8.\*
- 105: em\_gss9.\*
- 106: em\_gss10.\*
- 107: em\_gss11.\*
- 108: gb\_gss10.\*
- 109: gb\_gss11.\*
- 110: em\_gss12.\*
- 111: gb\_gss12.\*
- 112: gb\_gss13.\*
- 113: gb\_gss14.\*
- 114: gb\_gss15.\*
- 115: gb\_gss16.\*
- 116: gb\_gss17.\*

Result No.	Score	Query Match	Length	DB ID	Description
1	48	4.8	203	27	BB121888
2	48	4.8	239	28	BB243114
3	48	4.8	276	30	BB367625
4	48	4.8	283	28	BB229235
5	48	4.8	287	28	BB243023
6	48	4.8	560	114	A2038592
7	47	4.7	417	15	AU040346
8	47	4.7	537	113	AQ999562
9	46	4.6	163	32	BB549783
10	46	4.6	187	114	A2060889
11	46	4.6	199	117	A2229339
12	46	4.6	203	26	BB030960
13	46	4.6	256	22	AW532415
14	45	4.6	268	36	C78686
15	45	4.6	283	31	BB477575
16	45	4.6	295	15	AV077518
17	46	4.6	312	11	A1596969
18	46	4.6	312	30	BB392273
19	46	4.6	316	28	BB220457
20	46	4.6	317	11	A1515851
21	46	4.6	330	5	AA672427
22	46	4.6	330	32	BB549215
23	46	4.6	336	117	AZ286008
24	46	4.6	372	7	AA921447
25	46	4.6	391	114	AZ064788
26	46	4.6	399	1	AA067316
27	46	4.6	404	14	AU023235
28	46	4.6	410	8	A1012122
29	46	4.6	428	13	A1851901
30	46	4.6	430	11	A1553552
31	46	4.6	432	11	A1507388
32	46	4.6	436	9	A1178310
33	46	4.6	454	24	AW823066
34	46	4.6	457	10	A1412379
35	46	4.6	462	9	A1237187
36	46	4.6	466	33	BE105955
37	46	4.6	467	33	BE106153
38	46	4.6	475	8	A1013939
39	46	4.6	486	113	AZ015088
40	46	4.6	498	6	AA800808
41	46	4.6	503	6	AA851877
42	46	4.6	509	113	AQ979423
43	46	4.6	516	114	AZ067703
44	46	4.6	530	112	AQ920947
45	46	4.6	545	33	BE118632

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BB121888 203 bp mRNA EST 27-JUN-2000  
BB121888 RIKEN full-length enriched, adult male urinary bladder Mus musculus cDNA clone 9530083N07 3', mRNA sequence.  
BB121888  
BB121888  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 203)  
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,I., Ishikawa,J., Ishikawa,T., Itoh,M., Iizawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya,T., Tsunoda,Y., Watanabe,S., Yamamura,T., Yamana,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Genome Exploration Research Group, Life Science Tsukuba Center,  
Genome Science Laboratory  
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: +81-298-36-9013  
Fax: +81-298-36-9098  
Email: genome-res@rtc.riken.go.jp,  
URL: http://genome.rtc.riken.go.jp/  
Carninci,P., Nishiyama,Y., Westover,M. and Hayashizaki,Y.  
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES  
source

Location/Qualifiers  
1..203  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="RIKEN full-length enriched, adult male urinary bladder"  
/sex="male"  
/tissue\_type="urinary bladder"  
/lab\_host="DH10B"  
/dev\_stage="adult"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'  
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second

## ALIGNMENTS

RESULT  
BB121888

strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTAAATTAATCCCGCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 54 a 55 c 38 g 56 t  
ORIGIN

Query Match 4.8%; Score 48; DB 27; Length 203;  
Best Local Similarity 100.0%; Pred. No. 9.7e-14;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 aggcattcgatccattacagatggtgtgagccaccatgtgtttc 318  
|||||  
DB 67 AGGCATCGGATCCCATACAGATGTTGTGAGCCACCATGTGTTTC 114

RESULT 2  
BB243114 239 bp mRNA EST 06-JUL-2000  
LOCUS BB243114 RIKEN full-length enriched, 3 days neonate thymus Mus  
DEFINITION musculus cDNA clone A630094P22 3', mRNA sequence.

ACCESSION BB243114  
VERSION BB243114.1 GI:8935860  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
AUTHORS Konno, H., Alizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Genome Exploration Research Group, Life Science Tsukuba Center,  
Genome Science Laboratory  
The Institute of Physical and Chemical Research (RIKEN), Genomic  
Sciences Center Tsukuba, Ibaraki 305-0074, Japan  
Tel: +81-298-36-9013  
Fax: +81-298-36-9098  
Email: genome-res@rtc.riken.go.jp,  
URL: http://genome.rtc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.  
Location/Qualifiers  
1. .239  
/organism="Mus musculus"

/db\_xref="taxon:10090"  
/clone="A630094P22"  
/clone\_lib="RIKEN full-length enriched, 3 days neonate thymus"  
/tissue\_type="thymus"  
/dev\_stage="3 days neonate"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATTCGAGTAAATTAATCCCGCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 66 a 63 c 39 g 71 t  
ORIGIN

Query Match 4.8%; Score 48; DB 28; Length 239;  
Best Local Similarity 100.0%; Pred. No. 9.8e-14;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 aggcattcgatccattacagatggtgtgagccaccatgtgtttc 318  
|||||  
DB 103 AGGCATCGGATCCCATACAGATGTTGTGAGCCACCATGTGTTTC 150

RESULT 3

BB367625 276 bp mRNA EST 12-JUL-2000  
LOCUS BB367625 RIKEN full-length enriched, 16 days embryo head Mus  
DEFINITION musculus cDNA clone C130038M12 3', mRNA sequence.

ACCESSION BB367625  
VERSION BB367625.1 GI:9079454  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 276)

REFERENCE

AUTHORS

Konno, H., Alizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al.)

TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki  
Genome Exploration Research Group, Life Science Tsukuba Center,  
Genome Science Laboratory  
The Institute of Physical and Chemical Research (RIKEN), Genomic  
Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: +81-298-36-9013  
Fax: +81-298-36-9098  
Email: genome-res@rtc.riken.go.jp,





Db 147 AGGCATCGGATCCATTACAGATGGTTGTGAGCCACCATGTGTTTC 194

## RESULT 5

BB243023 287 bp mRNA EST 06-JUL-2000  
 LOCUS BB243023 RIKEN full-length enriched, 3 days neonate thymus Mus  
 DEFINITION musculus cDNA clone A630094J20 3', mRNA sequence.  
 BB243023  
 ACCESSION BB243023.1 GI:8935769  
 VERSION EST.  
 KEYWORDS house mouse.  
 SOURCE house musculus

## ORGANISM

Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 287)  
 Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
 P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
 Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,  
 Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,  
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,  
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
 Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata  
 Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suganara,Y.,  
 Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya  
 T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,  
 Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino  
 M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Konno,H., et al.)

## TITLE

## JOURNAL

## COMMENT

Unpublished (2000)  
 Contact: Yoshihide Hayashizaki  
 Genome Exploration Research Group, Life Science Tsukuba Center,  
 Genome Science Laboratory  
 The Institute of Physical and Chemical Research (RIKEN), Genomic  
 Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: +81-298-36-9013  
 Fax: +81-298-36-9098  
 Email: genome-res@tc.riken.go.jp  
 URL: http://genome.ttc.riken.go.jp/  
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Thermostabilization and thermoactivation of thermostable enzymes by  
 trehalose and its application for the synthesis of full length  
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
 Y. and Hayashizaki,Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (http://genome.ttc.riken.go.jp) for  
 further details.

## FEATURES

## SOURCE

1. 287  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="A630094J20"  
 /clone\_lib="RIKEN full-length enriched, 3 days neonate  
 thymus"  
 /tissue="thymus"  
 /dev\_stage="3 days neonate"  
 /lab\_host="DH10B"  
 /note="Site 1: Sali; Site 2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGAGGATCCCAAGAGCTCTTTTCTTTTCTTTTCTTTTCTTTN 3'] cDNA was

prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. cDNA went through one round of normalization  
 to Rot -20.0 and subtraction to Rot -459.0. Second  
 strand cDNA was prepared with the primer adapter of  
 sequence [5' GAGAGAGATTCGAGTTAATTAATATCCCCCCCCC  
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
 modified pluescript KS(+) after bulk excision from Lambda  
 FLC I."

BASE COUNT 97 a 60 c 54 g 76 t  
 ORIGIN

Query Match 4.8%; Score 48; DB 28; Length 287;  
 Best Local Similarity 100.0%; Pred. No. 1e-13;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 agggcgcggtccattacagatggtgtgagccaccatgtggttc 318

|||||  
 Db 151 AGGCATCGGATCCATTACAGATGGTTGTGAGCCACCATGTGTTTC 198

## RESULT 6

AZ038592  
 LOCUS RPCI-23-378A20.TV RPCI-23 Mus musculus genomic clone RPCI-23-378A20  
 DEFINITION DNA sequence.  
 AZ038592 560 bp DNA GSS 01-MAR-2000  
 RPCI-23-378A20.TV RPCI-23 Mus musculus genomic clone RPCI-23-378A20

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## Other\_GSSs

## Contact

## Department of Eukaryotic Genomics

## The Institute for Genomic Research

## 9712 Medical Center Dr., Rockville, MD 20850, USA

## Tel: 301 838 0200

## Fax: 301 838 0208

## Email: szhao@tigr.org

## Clones are derived from the mouse BAC library RPCI-23. For BAC

## library availability, please contact Pieter de Jong

## (pieter@dejong.med.buffalo.edu). Clones may be purchased from

## BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

## or from Resea ch Genetics (info@resgen.com). BAC end page:

## http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html

## Plate: 378 row: A column: 20

## Seq primer: T7

## Class: BAC ends.

## Location/Qualifiers

## 1. .560

## /organism="Mus musculus"

## /strain="C57BL/6J"

## /db\_xref="taxon:10090"

## /clone="RPCI-23-378A20"

## /clone\_lib="RPCI-23"

## /sex="Female"

## /lab\_host="DH10B"

## /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site\_1:

## EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or

## brain genomic DNA was isolated and partially digested

## with a combination of EcoRI and EcoRI Methylase. Size

## selected DNA was cloned into the pBAC3.6 vector at the

## EcoRI sites. The ligation products were transformed into

## DH10B electrocompetent cells (BRL Life Technologies)."

```

BASE COUNT      141 a 138 c 133 g 148 t
ORIGIN

Query Match      4.8%; Score 48; DB 114; Length 560;
Best Local Similarity 100.0%; Pred. No. 1e-13;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 agggcattgacccattacagatggttgagccacacatgtgtttc 318
Db 387 AGGGATCGGATCCCATACAGATGGTTGTGAGCCACCATGTGTTTC 434

RESULT 7
LOCUS      AU040346      417 bp      mRNA      EST      04-DEC-1998
DEFINITION      AU040346 Mouse four-cell-embryo cDNA Mus musculus cDNA clone
VERSION      J0804E08 3', mRNA sequence.
ACCESSION      AU040346
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 417)
AUTHORS      KO, M.S.H., Kitchen, J.R., Wang, X., Threat, T.A., Sun, T., DePalma, G.E.,
Liang, Y., Kargul, G.Y., Sharara, R., Lim, M.K., and Doi, H.
TITLE      Systematic analyses of genes expressed in 4-cell mouse embryo (The
JOURNAL      ERATO/Doi Project at Wayne State University)
COMMENT      Unpublished (1998)
Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
Wag Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdoe@jst.go.jp.
FEATURES
source
1. .417
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0804E08"
/dev_stage="four-cell-embryo"
BASE COUNT      114 a 95 c 86 g 122 t
ORIGIN

Query Match      4.7%; Score 47; DB 15; Length 417;
Best Local Similarity 100.0%; Pred. No. 3.4e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 caggaattgaactcagacatttggaagcagtcagtcgtcttaac 394
Db 261 CAGGAATTGACTCAGGACCTTTTGAAGAGCAGTCAGTCTCTTAAC 307

RESULT 8
LOCUS      AQ999562      537 bp      DNA      GSS      24-FEB-2000
DEFINITION      RPCI-23-387K11.TV RPCI-23 Mus musculus genomic clone RPCI-23-387K11
, DNA sequence.
ACCESSION      AQ999562
VERSION      AQ999562.1 GI:7074659
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 537)
AUTHORS      Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akinret,
B., Levins, M., McGann, S., Tseng, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-387K11.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301-838 0200
Fax: 301-838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pleter@tigr.org, med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 387 row: K column: 11
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
1. .537
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-387K11"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
ECORI; Site_2: ECORI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      150 a 135 c 94 g 158 t
ORIGIN

Query Match      4.7%; Score 47; DB 113; Length 537;
Best Local Similarity 100.0%; Pred. No. 3.4e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 agggcattgacccattacagatggttgagccacacatgtgttt 317
Db 116 AGGGATCGGATCCCATACAGATGGTTGTGAGCCACCATGTGTTT 162

RESULT 9
LOCUS      BS549783      163 bp      mRNA      EST      31-JUL-2000
DEFINITION      BS549783 RIKEN full-length enriched, 2 days pregnant adult female
oviduct Mus musculus cDNA clone E230017122 3', mRNA sequence.
ACCESSION      BS549783
VERSION      BS549783.1 GI:9621211
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 163)
AUTHORS      Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, K., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
Y., Shimamoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, F., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
M., Murumatsu, M. and Hayashizaki, Y.

```





UI-R-BS0-amo-b-10-0-UI 3', mRNA sequence.

ACCESSION  
AW532415  
VERSION  
AW532415.1 GI:7174829  
KEYWORDS  
EST.  
SOURCE  
Norway rat.

ORGANISM  
Rattus norvegicus

REFERENCE  
1 (bases 1 to 256)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
non-normalized embryo at 13 dpc library cDNA library preparation:  
M.B. Soares Lab Clone distribution: clones will be available  
through Research Genetics (www.resgen.com) The following repetitive  
elements were found in this cDNA sequence: 1-192, >B2#SINE/B2  
198-256, >ORR1B#LTR/MaLR  
Seq primer: M13 Forward  
POLYA-Yes

FEATURES

source

Location/Qualifiers

1..256

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-BS0-amo-b-10-0-UI"

/clone\_lib="UI-R-BS0"

/dev\_stage="embryonic 13 dpc"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site1: Not I; Site2: Eco RI; The UI-R-BS0  
library is derived from 13 dpc whole embryo tissue. For a  
detailed description of the library from which this clone  
was derived, please visit our web site at  
ratest.eng.uiowa.edu.  
TAG\_LIB=UI-R-BS0  
TAG\_TISSUE=embryo at 13 dpc  
TAG\_SEQ=AAATCC"

BASE COUNT 63 a 55 c 56 g 82 t

ORIGIN

Query Match 4.6%; Score 46; DB 22; Length 256;

Best Local Similarity 100.0%; Pred. No. 1.1e-12;

Mismatches 0; Indels 0; Gaps 0;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 agggcgcgattccattacagattgtgtgagccaccattgtggtt 316

|||||

Db 75 AGGCGATCGATCCATTACAGATGTTGTGAGCCACCATGTGGTT 120

|||||

RESULT 14

C78686

LOCUS

DEFINITION

J0053C06 3' similar to R.norvegicus gene encoding multifunctional,  
mRNA sequence.

ACCESION

VERSION

C78686.1 GI:2979445

25-JUN-1998

EST

CDNA clone

musculus

gene

encoding multifunctional,

mRNA sequence.

C78686

C78686.1

GI:2979445

KEYWORDS

SOURCE

ORGANISM

house mcuse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 268)

Ko,M.S.F., Kitchen,J.R., Wang,X., Wang,X., Threat,T.A., Sun,T.,  
Grahovac,M.J., Mason,R., Lim,M.K., Paonessa,P.D., Sauls,A.D. and  
Doi,H.

TITLE

Systematic analyses of genes expressed in 3.5-dpc mouse blastocyst  
(The ERATO/Doi Project at Wayne State University)

Unpublished (1997)

On Oct 17, 1997 this sequence version replaced gi:2519016.

COMMENT

Contact: Hirofumi Doi

Doi Biosymmetry Project, ERATO

Japan Science and Technology Corporation (JST)

WBG Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan

Email: fdoioa.jst.go.jp.

FEATURES

source

Location/Qualifiers

1..268

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="J0053C06"

/clone\_lib="Mouse 3.5-dpc blastocyst cDNA"

/tissue\_type="blastocyst"

/dev\_stage="3.5-dpc"

BASE COUNT 67 a 73 c 46 g 82 t

ORIGIN

Query Match 4.6%; Score 46; DB 36; Length 268;

Best Local Similarity 100.0%; Pred. No. 1.1e-12;

Mismatches 0; Indels 0; Gaps 0;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 ggaattgactcaggaccttggagcagtcagtcagtcgtcttaacc 395

|||||

Db 111 GGAATTGAATCAGGACCTTTGGAGAGCAGTCAGTCGTCTTAACC 156

|||||

RESULT 15

BB477575/c

LOCUS

DEFINITION

BB477575 RIKEN full-length enriched, 13 days embryo heart Mus

musculus CDNA clone D330025K24 3', mRNA sequence.

ACCESION

BB477575

VERSION

BB477575.1 GI:9395184

KEYWORDS

house mcuse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 283)

Kanno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci

,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,

Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,

Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,

Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.

,Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.

,Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata

,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.

,Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya

,T., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.

,Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino

,M., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mice ESTs (Kanno,H., et al.)

Unpublished (2000)

CONTACT: Yoshihide Hayashizaki

Genome Exploration Research Group, Life Science Tsukuba Center,

Genome Science Laboratory

The Institute of Physical and Chemical Research (RIKEN), Genomic

Sciences-Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan



us-09-101-423a-3.rge

Mon Nov 6 10:14:20 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 4, 2000, 00:36:55 ; Search time 7109.34 Seconds  
(without alignments)  
619.259 Million cell updates/sec

Title: US-09-101-423A-3  
Perfect score: 1008  
Sequence: 1 aggaccagagttcacatccc.....ttctgtgtctctctggaac 1008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues 2067340  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba1:  
2: gb\_ba2:  
3: gb\_om:  
4: gb\_ov:  
5: gb\_pat:  
6: gb\_ph:  
7: gb\_pl1:  
8: gb\_pl2:  
9: gb\_pr1:  
10: gb\_pr2:  
11: gb\_pr3:  
12: em\_fun:  
13: em\_hum1:  
14: em\_hum2:  
15: em\_in:  
16: em\_om:  
17: em\_or:  
18: em\_ov:  
19: em\_pat:  
20: em\_ph:  
21: em\_pl:  
22: em\_to:  
23: em\_sts:  
24: em\_sy:  
25: em\_un:  
26: em\_vi:  
27: gb\_htg1:  
28: gb\_htg2:  
29: gb\_in1:  
30: gb\_in2:  
31: gb\_ba1:  
32: gb\_ba2:  
33: em\_hum3:  
34: em\_hum4:  
35: gb\_pr4:  
36: gb\_htg3:  
37: gb\_htg4:  
38: gb\_htg5:  
39: gb\_htg6:  
40: gb\_htg7:  
41: em\_htg1:  
42: em\_htg2:  
43: em\_htg3:  
44: em\_hum5:  
45: gb\_pl3:  
46: gb\_pr5:  
47: gb\_htg8:  
48: gb\_htg9:  
49: gb\_htg10:  
50: gb\_htg11:  
51: gb\_htg12:  
52: gb\_htg13:  
53: gb\_htg14:  
54: gb\_in3:  
55: gb\_htg15:  
56: gb\_htg16:  
57: gb\_htg17:  
58: em\_htg4:  
59: em\_htg5:  
60: em\_htg6:  
61: em\_htg7:  
62: em\_hum6:  
63: gb\_htg18:  
64: gb\_htg19:  
65: gb\_ba3:  
66: em\_htg8:  
67: em\_htg9:  
68: em\_htg10:  
69: em\_htg11:  
70: em\_htg12:  
71: em\_htg13:  
72: em\_htg14:  
73: em\_htg15:  
74: em\_htg16:  
75: em\_htg17:  
76: em\_htg18:  
77: em\_htg19:  
78: em\_htg20:  
79: em\_htg21:  
80: em\_htg22:  
81: em\_htg23:  
82: gb\_pr6:  
83: gb\_pr7:  
84: gb\_htg20:  
85: gb\_htg21:  
86: gb\_htg22:  
87: gb\_htg23:  
88: gb\_ro:  
89: gb\_sts1:  
90: gb\_sts2:  
91: gb\_sy:  
92: gb\_un:  
93: gb\_vil:  
94: gb\_vil2:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1008	100.0	1008	5	A64060	A64060 Sequence 3
2	142.4	14.1	201251	50	AC025964	AC025964 Mus muscu
3	140.8	14.0	207737	49	AC023934	AC023934 Homo sapi
4	135.6	13.5	166354	51	AC026380	AC026380 Mus muscu
5	134.6	13.4	267762	57	AC073817	AC073817 Mus muscu
6	131.2	13.0	193260	50	AC025598	AC025598 Homo sapi
7	130.2	12.9	205386	56	AC069465	AC069465 Mus muscu
8	130	12.9	236195	57	AC073713	AC073713 Mus muscu
9	129.6	12.9	157207	49	AC023802	AC023802 Mus muscu
10	127	12.6	242733	50	AC025581	AC025581 Mus muscu
11	124.4	12.3	213741	48	AC022775	AC022775 Mus muscu
12	123.4	12.2	194162	51	AC027158	AC027158 Homo sapi

## ALIGNMENTS

Query Match	100.0%	Score 1008;	DB 5;	Length 1008;
Best Local Similarity	100.0%;	Pred. No. 2.7e-183;		
Matches 1008;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

	QY	DB	QY	DB	0; Gaps	0; Indels	0; Mismatches
1	aggaccagaggttcacatccatcaaatggccagaaggttttaatgctgtcttttggccc	60	1	AGGACCAGAGTTTACATTCCTCATTAATGGCCAGAGGTTTTTAATGCTGCTTTTGGCCC	60		
61	agggggcaactgcacacacatgtgcacatacacttcacagacacacattcgcgcagcata	120	61	AGGGGGCAAGTCGACACACATGTGCACATACACTTCAGAGACACACATTCAGCAGCATA	120		

LOCUS	AC025964	201251 bp	DNA	HTG	09-JUL-2000
DEFINITION	Mus musculus chromosome 11 clone RP23-136C7 map 11, WORKING DRAFT				
ACCESSION	SEQUENCE: 48 unordered pieces.				
VERSION	AC025964				
KEYWORDS	AC025964.2 GI:8980918				
SOURCE	HTG; HTG3_PHASE1; HTG3_DRAFT.				
ORGANISM	house mouse.				
	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus;				

RESULT 2  
AC025964/C  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 201251)  
Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Mus musculus chromosome 11, clone RP23-136C7  
Unpublished  
2 (bases 1 to 201251)  
Biren,B., Linton,L., Nusbaum,C., Lander,E., Brown,A., Burkett,G.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Choepe,Y., Colangelo,M., Collins,S.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Glnde,S., Govette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheeters,R.,  
Meidlm,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

Submitted (18-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 9, 2000 this sequence version replaced gi:7262752.

COMMENT

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L5729  
Center clone name: 136\_C7

----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 178807 bases at least Q40  
Consensus quality: 191016 bases at least Q30  
Consensus quality: 194940 bases at least Q20  
Insert size: 173000; agarose-fp  
Insert size: 196551; sum-of-contigs  
Quality coverage: 3.7 in Q20 bases; agarose-fp  
Quality coverage: 3.2 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 48 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1487: contig of 1487 bp in length  
\* 1488 1587: gap of 100 bp  
\* 1588 2793: contig of 1206 bp in length  
\* 2794 2893: gap of 100 bp  
\* 2894 4247: contig of 1354 bp in length  
\* 4248 4347: gap of 100 bp  
\* 4348 5849: contig of 1502 bp in length  
\* 5850 5949: gap of 100 bp  
\* 5950 7248: contig of 1299 bp in length  
\* 7249 7348: gap of 100 bp  
\* 7349 8876: contig of 1528 bp in length  
\* 8877 8976: gap of 100 bp  
\* 10439: contig of 1463 bp in length  
\* 10440 10539: gap of 100 bp  
\* 10540 12180: contig of 1541 bp in length  
\* 12181 14702: contig of 2522 bp in length  
\* 14703 14802: gap of 100 bp  
\* 14803 16733: contig of 1931 bp in length  
\* 16734 16833: gap of 100 bp  
\* 16834 18755: contig of 1922 bp in length  
\* 18756 18955: gap of 100 bp  
\* 18956 20786: contig of 1931 bp in length  
\* 20787 20886: gap of 100 bp  
\* 20887 23125: contig of 2239 bp in length  
\* 23126 23225: gap of 100 bp  
\* 23226 25267: contig of 2042 bp in length  
\* 25268 25367: gap of 100 bp  
\* 25368 27294: contig of 1927 bp in length  
\* 27295 27394: gap of 100 bp  
\* 27395 29083: contig of 1689 bp in length  
\* 29084 29183: gap of 100 bp  
\* 29184 31810: contig of 2627 bp in length  
\* 31811 31910: gap of 100 bp  
\* 31911 34614: contig of 2704 bp in length  
\* 34615 34714: gap of 100 bp  
\* 34715 36824: contig of 2110 bp in length  
\* 36825 36924: gap of 100 bp  
\* 36925 39358: contig of 2434 bp in length  
\* 39359 39458: gap of 100 bp  
\* 39459 41675: contig of 2217 bp in length  
\* 41676 41775: gap of 100 bp  
\* 41776 44142: contig of 2367 bp in length  
\* 44143 44242: gap of 100 bp  
\* 44243 45375: contig of 1133 bp in length  
\* 45376 45475: gap of 100 bp  
\* 45476 48030: contig of 2555 bp in length  
\* 48031 48130: gap of 100 bp  
\* 48131 50762: contig of 2632 bp in length  
\* 50763 50862: gap of 100 bp  
\* 50863 54865: contig of 4003 bp in length  
\* 54866 54965: gap of 100 bp  
\* 54966 58786: contig of 3821 bp in length  
\* 58787 58886: gap of 100 bp  
\* 58887 62079: contig of 3193 bp in length  
\* 62080 62179: gap of 100 bp  
\* 62180 66768: contig of 4589 bp in length  
\* 66769 66868: gap of 100 bp  
\* 66869 70876: contig of 4008 bp in length  
\* 70877 70976: gap of 100 bp  
\* 70977 75996: contig of 5020 bp in length  
\* 75997 76096: gap of 100 bp  
\* 76097 80273: contig of 4177 bp in length  
\* 80274 80373: gap of 100 bp  
\* 80374 85679: contig of 5306 bp in length  
\* 85680 85779: gap of 100 bp  
\* 85780 91717: contig of 5938 bp in length  
\* 91718 91817: gap of 100 bp  
\* 91818 97046: contig of 5229 bp in length  
\* 97047 97146: gap of 100 bp  
\* 97147 101127: contig of 3981 bp in length  
\* 101128 101227: gap of 100 bp  
\* 101228 106788: contig of 5561 bp in length  
\* 106789 106888: gap of 100 bp  
\* 106889 111554: contig of 4666 bp in length  
\* 111555 111654: gap of 100 bp  
\* 111655 118598: contig of 6944 bp in length  
\* 118599 118698: gap of 100 bp  
\* 118699 124207: contig of 5509 bp in length  
\* 124208 124307: gap of 100 bp  
\* 124308 130324: contig of 6017 bp in length  
\* 130325 130424: gap of 100 bp  
\* 130425 138528: contig of 8104 bp in length  
\* 138529 138628: gap of 100 bp  
\* 138629 146991: contig of 8363 bp in length



Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: 15621  
 Center clone name: 618\_P\_13

----- Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 185047 bases at least Q40  
 Consensus quality: 196210 bases at least Q30  
 Consensus quality: 200872 bases at least Q20  
 Insert size: 157000; agarose-fp  
 Insert size: 204357; sum-of-contigs  
 Quality coverage: 4.4 in Q20 bases; agarose-fp  
 Quality coverage: 3.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1178: contig of 1178 bp in length
* 1179 1278: gap of 100 bp
* 1279 1431: contig of 153 bp in length
* 1432 1531: gap of 100 bp
* 1532 2672: contig of 1141 bp in length
* 2673 2772: gap of 100 bp
* 2773 3953: contig of 1081 bp in length
* 3954 5494: contig of 1541 bp in length
* 5495 5594: gap of 100 bp
* 5595 6808: contig of 1214 bp in length
* 6809 6908: gap of 100 bp
* 6909 8821: contig of 1913 bp in length
* 8822 8921: gap of 100 bp
* 8922 10227: contig of 1306 bp in length
* 10228 10327: gap of 100 bp
* 10328 12417: contig of 2090 bp in length
* 12418 12517: gap of 100 bp
* 12518 14413: contig of 1896 bp in length
* 14414 14513: gap of 100 bp
* 14514 16780: contig of 2267 bp in length
* 16781 16880: gap of 100 bp
* 16881 18989: contig of 2109 bp in length
* 18990 19089: gap of 100 bp
* 19090 21339: contig of 2250 bp in length
* 21340 21439: gap of 100 bp
* 21440 24496: contig of 3057 bp in length
* 24497 24596: gap of 100 bp
* 24597 27488: contig of 2892 bp in length
* 27489 27588: gap of 100 bp
* 27589 30111: contig of 2523 bp in length
* 30112 30211: gap of 100 bp
* 30212 33587: contig of 3376 bp in length
* 33588 33687: gap of 100 bp
* 33688 37578: contig of 3891 bp in length
* 37579 37678: gap of 100 bp
* 37679 42795: contig of 5117 bp in length
* 42796 42895: gap of 100 bp
* 42896 48285: contig of 5390 bp in length
* 48286 48385: gap of 100 bp
* 48386 52171: contig of 3786 bp in length
* 52172 52271: gap of 100 bp
* 52272 60414: contig of 8143 bp in length
* 60415 60514: gap of 100 bp
* 60515 68386: contig of 7872 bp in length
* 68387 68486: gap of 100 bp
* 68487 75220: contig of 6734 bp in length
* 75221 75320: gap of 100 bp

```

```

* 75321 82520: contig of 7200 bp in length
* 82521 82620: gap of 100 bp
* 82621 92012: contig of 9392 bp in length
* 92013 92112: gap of 100 bp
* 92113 100585: contig of 8473 bp in length
* 100586 100685: gap of 100 bp
* 100686 111797: contig of 11112 bp in length
* 111798 111897: gap of 100 bp
* 111898 120279: contig of 8382 bp in length
* 120280 120379: gap of 100 bp
* 120380 132436: contig of 12057 bp in length
* 132437 132536: gap of 100 bp
* 132537 144076: contig of 11540 bp in length
* 144077 144176: gap of 100 bp
* 144177 160302: contig of 16126 bp in length
* 160303 160402: gap of 100 bp
* 160403 175068: contig of 14666 bp in length
* 175069 175168: gap of 100 bp
* 175169 192239: contig of 17071 bp in length
* 192240 192339: gap of 100 bp
* 192340 207757: contig of 15418 bp in length.

```

#### FEATURES

```

source
1..207757
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="17"
  /map="17"
  /clone_lib="RPC1-11 Human Male BAC"
  /clone="RP11-618P13"
  /note="assembly_fragment"
  1..1178
  1279..1431
  /note="assembly_fragment"
  /note="assembly_fragment"
  clone_end:SP6
  vector_side:right
  1532..2672
  /note="assembly_fragment"
  2773..3853
  /note="assembly_fragment"
  3954..5494
  /note="assembly_fragment"
  5595..6808
  /note="assembly_fragment"
  6909..8821
  /note="assembly_fragment"
  8922..10227
  /note="assembly_fragment"
  10328..12417
  /note="assembly_fragment"
  12518..14413
  /note="assembly_fragment"
  14514..16780
  /note="assembly_fragment"
  16881..18989
  /note="assembly_fragment"
  19090..21339
  /note="assembly_fragment"
  21440..24496
  /note="assembly_fragment"
  24597..27488
  /note="assembly_fragment"
  27589..30111
  /note="assembly_fragment"
  30212..33587
  /note="assembly_fragment"
  33688..37578
  /note="assembly_fragment"
  37679..42795
  /note="assembly_fragment"
  42896..48285
  /note="assembly_fragment"
  48386..52171
  /note="assembly_fragment"

```





















Consensus quality: 197506 bases at least Q30  
Consensus quality: 204301 bases at least Q20  
Insert size: 209241; sum-of-contigs  
Quality coverage: 3.4 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 46 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 223 322: contig of 222 bp in length  
\* 223 322: gap of 100 bp  
\* 323 1808: contig of 1486 bp in length  
\* 1809 1908: gap of 100 bp  
\* 1909 3599: contig of 1691 bp in length  
\* 3600 3599: gap of 100 bp  
\* 3700 4899: contig of 1200 bp in length  
\* 4900 4999: gap of 100 bp  
\* 5000 6752: contig of 1753 bp in length  
\* 6753 6852: gap of 100 bp  
\* 6853 7903: contig of 1051 bp in length  
\* 7904 8003: gap of 100 bp  
\* 8004 9229: contig of 1226 bp in length  
\* 9230 9329: gap of 100 bp  
\* 9330 9949: contig of 620 bp in length  
\* 9950 10049: gap of 100 bp  
\* 10050 12000: contig of 1951 bp in length  
\* 12001 12100: gap of 100 bp  
\* 12101 14173: contig of 2073 bp in length  
\* 14174 14273: gap of 100 bp  
\* 14274 15830: contig of 1557 bp in length  
\* 15831 15930: gap of 100 bp  
\* 15931 17552: contig of 1622 bp in length  
\* 17553 17652: gap of 100 bp  
\* 17653 19933: contig of 2281 bp in length  
\* 19934 20033: gap of 100 bp  
\* 20034 22678: contig of 2645 bp in length  
\* 22679 22778: gap of 100 bp  
\* 22779 24995: contig of 2217 bp in length  
\* 24996 25095: gap of 100 bp  
\* 25096 29357: contig of 4262 bp in length  
\* 29358 29457: gap of 100 bp  
\* 29458 32930: contig of 3473 bp in length  
\* 32931 33030: gap of 100 bp  
\* 33031 35864: contig of 2834 bp in length  
\* 35865 35964: gap of 100 bp  
\* 35965 38448: contig of 2484 bp in length  
\* 38449 38548: gap of 100 bp  
\* 38549 41879: contig of 3331 bp in length  
\* 41880 41979: gap of 100 bp  
\* 41980 44270: contig of 2291 bp in length  
\* 44271 44370: gap of 100 bp  
\* 44371 48637: contig of 4267 bp in length  
\* 48638 48737: gap of 100 bp  
\* 48738 53337: contig of 4600 bp in length  
\* 53338 53437: gap of 100 bp  
\* 53438 57180: contig of 3743 bp in length  
\* 57181 57280: gap of 100 bp  
\* 57281 59790: contig of 2510 bp in length  
\* 59791 59890: gap of 100 bp  
\* 59891 64641: contig of 4751 bp in length  
\* 64642 64741: gap of 100 bp  
\* 64742 69188: contig of 4447 bp in length  
\* 69189 69288: gap of 100 bp  
\* 69289 72951: contig of 3663 bp in length  
\* 72952 73051: gap of 100 bp  
\* 73052 78212: contig of 5161 bp in length  
\* 78213 78312: gap of 100 bp  
\* 78313 82821: contig of 4509 bp in length  
\* 82822 82921: gap of 100 bp

\* 82922 88525: contig of 5604 bp in length  
\* 88526 88625: gap of 100 bp  
\* 88626 93132: contig of 4507 bp in length  
\* 93133 93232: gap of 100 bp  
\* 93233 99012: contig of 5780 bp in length  
\* 99013 99112: gap of 100 bp  
\* 99113 105347: contig of 6235 bp in length  
\* 105348 105447: gap of 100 bp  
\* 105448 110428: contig of 4981 bp in length  
\* 110429 110528: gap of 100 bp  
\* 110529 115562: contig of 5034 bp in length  
\* 115563 115662: gap of 100 bp  
\* 115663 122718: contig of 7056 bp in length  
\* 122719 122818: gap of 100 bp  
\* 122819 128623: contig of 5805 bp in length  
\* 128624 128723: gap of 100 bp  
\* 128724 138514: contig of 9791 bp in length  
\* 138515 138614: gap of 100 bp  
\* 138615 145431: contig of 6817 bp in length  
\* 145432 145531: gap of 100 bp  
\* 145532 150633: contig of 5102 bp in length  
\* 150634 150733: gap of 100 bp  
\* 150734 159378: contig of 8645 bp in length  
\* 159379 159478: gap of 100 bp  
\* 159479 169395: contig of 9917 bp in length  
\* 169396 169495: gap of 100 bp  
\* 169496 181416: contig of 11921 bp in length  
\* 181417 181516: gap of 100 bp  
\* 181517 195657: contig of 14141 bp in length  
\* 195658 195757: gap of 100 bp  
\* 195758 213741: contig of 17984 bp in length.

FEATURES  
Location/Qualifiers  
1. .213741  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-71017"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
misc\_feature  
1. .222  
/note="assembly\_fragment"  
clone\_end:77  
vector\_side:right  
misc\_feature  
323. .1808  
/note="assembly\_fragment"  
misc\_feature  
1909. .3599  
/note="assembly\_fragment"  
misc\_feature  
3700. .4899  
/note="assembly\_fragment"  
misc\_feature  
5000. .6752  
/note="assembly\_fragment"  
misc\_feature  
6853. .7903  
/note="assembly\_fragment"  
misc\_feature  
8004. .9229  
/note="assembly\_fragment"  
misc\_feature  
9330. .9949  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
misc\_feature  
10050. .12000  
/note="assembly\_fragment"  
misc\_feature  
12101. .14173  
/note="assembly\_fragment"  
misc\_feature  
14274. .15830  
/note="assembly\_fragment"  
misc\_feature  
15931. .17552  
/note="assembly\_fragment"  
misc\_feature  
17653. .19933  
/note="assembly\_fragment"  
misc\_feature  
20034. .22678  
/note="assembly\_fragment"  
misc\_feature  
22779. .24995  
/note="assembly\_fragment"  
misc\_feature  
25096. .29357  
/note="assembly\_fragment"













Mon Nov 6 10:14:20 2000

us-09-101-423a-3.rge

---

Page 21



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:16:10 ; Search time 254.92 Seconds  
(without alignments)  
1485.439 Million cell updates/sec

Title: US-09-101-423a-3  
Perfect score: 1008  
Sequence: 1 aggaccagattcacatccc.....ttcctgtgtcctctggac 1008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*

1:	/SID56/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/SID56/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/SID56/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/SID56/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/SID56/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/SID56/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/SID56/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/SID56/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/SID56/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/SID56/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/SID56/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/SID56/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/SID56/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/SID56/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/SID56/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/SID56/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1001.6	99.4	1008	18	T72787
2	109.4	10.9	44576	21	Metastasis inducin
3	107.2	10.6	1210	16	Cosmid CVO14 conta
4	105	10.4	48974	20	Rat fatty acid bin
5	104.8	10.4	13011	18	Mouse Presenilin-1
6	104.4	10.4	90050	21	cDNA encoding rat
7	99	9.8	90050	21	Wild type (C57BL/6
8	97.4	9.7	5889	20	Wild type (C57BL/6
9	95.8	9.5	16956	21	Mouse A-myb genomi
10	92.8	9.2	49999	20	Wild type (C57BL/6
11	91.6	9.1	262	21	Murine LOBO homolo
12	91.2	9.0	48974	20	Rat cDNA clone rA5
					Mouse Presenilin-1

13	89	8.8	5109	17	T72654
14	89	8.8	6727	18	T88014
15	87.2	8.7	49999	20	223891
16	86.4	8.6	16009	20	223896
17	85.8	8.5	6480	19	X86357
18	85.2	8.5	44576	21	T99572
19	85	8.4	13206	11	261522
20	85	8.4	13206	11	005243
21	84.4	8.4	26698	17	T08126
22	84.4	8.4	26700	15	067902
23	84.4	8.4	26700	19	V15946
24	84.4	8.4	26700	20	V81283
25	80.8	8.0	37940	20	201026
26	80.2	8.0	178	20	X85695
27	79	7.8	49999	20	223891
28	77.8	7.7	49999	20	223896
29	77.4	7.7	2858	16	075270
30	77.4	7.7	13146	18	T96719
31	76.4	7.6	1072	20	210359
32	75.6	7.5	1104	20	210360
33	75.4	7.5	4263	20	V72672
34	75.4	7.5	51259	18	X83007
35	75.2	7.5	4580	17	T32034
36	75	7.4	233	20	X27002
37	73.4	7.3	5006	21	A29965
38	73	7.2	4989	16	085962
39	73	7.2	5889	20	V84328
40	72.6	7.2	5887	17	T65001
41	72.2	7.2	36901	20	223892
42	72.2	7.2	38886	20	223897
43	72	7.1	2294	16	092940
44	72	7.1	3519	16	092937
45	71.4	7.1	3217	20	Z00365

## ALIGNMENTS

RESULT 1	
T72787	
ID	T72787 standard; DNA; 1008 BP.
XX	
AC	T72787;
XX	
DT	22-SEP-1997 (first entry)
XX	
DE	Metastasis inducing DNA C6.
XX	
KW	Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;
KW	osteopontin; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9725443-A1.
XX	
PD	17-JUL-1997.
XX	
PF	10-JAN-1997; 97WO-GE000074.
XX	
PR	10-JAN-1996; 96GB-0000470.
XX	
PA	(UWLI-) UNIV LIVERPOOL.
XX	
PI	Barracough BR; Rudland PS;
XX	
DR	WPI; 1997-372878/34.
XX	
PT	New isolated metastasis-inducing DNA - used to develop products to
PT	identify and treat patients at risk from metastatic tumours
XX	
PS	Claim 10; Page 26; 38pp; English.
XX	
CC	Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12



Best Local Similarity 72.3%; Pred. NO. 1e-14;  
Matches 170; Conservative 0; Mismatches 36; Indels 29; Gaps 1;

Qy	199	ctttatttttttaaagatttattttaaagtatatagtagaacatgcctctccctccag	2358
Db	2389	ctttatttttttaaagatatatttatttatgtatgatgtcacttagctgtcttcag	2448
Qy	259	acatacagtagagggcatcgatccattacagatggtgtgagccacatggtgttc	318
Db	2449	acacaccagaagggggcatcggatcccatcacagatggctgagccaacatgctgtcgc	2508
Qy	319	acagatggtgtgtagccaccaatggtgttcaggattgaactcaggaccttggagaagc	378
Db	2509	-----tggggattgaacttaggacctctggaagagc	2539
Qy	379	agtcagtgtctttaaacctgaagccatctctctgaccttatatacaatttaa	433
Db	2540	aatcaatgctttaacgctgaccatctctcaacctgaaacttatattttaa	2594

RESULT 3  
Q87049/c  
ID Q87049 standard; DNA; 1210 BP.

AA 087049:

11-JAN-1996 (first entry)

XX Rat fatty acid binding protein (FABP) gene fragment.

XX Fatty acid binding protein; promoter; transgene; gastro-intestinal;  
KW ds.  
KW

Rattus sp.

XX  
PN  
WQ9511299-A1XX  
27-APR-1995XX  
PE 13-OCT-1994. 94WO-US11716

XX 19-00T-1003. 03MS-0141323

XX  
DA (AVCE-) AMCEN TNC

DEF	4000000	100	1-1-400	100
				XX

[illegible]

XX Enhanced trans:gene expression in specific tissues - of the  
PT gastrointestinal tract of a non-human animal  
PT

XX  
PS  
Claim 1: Page 38: 52pp: English.

Q87049 represents a DNA fragment containing the rat intestinal fatty  
acid binding protein (FABP) promoter, from position 1 and 1210 bp in  
the rat FABP gene. The FABP promoter sequence is used to construct  
an FABP promoter/interleukin-8 transgene which is ultimately  
inserted into a plasmid pF8-PA. The plasmid vector is then  
injected into mouse embryos and the transgene is expressed *in vivo*.  
Depending on the transgene(s) inserted into the animal and the  
level of transgene expression produced, the animal may become  
either increasingly susceptible or less susceptible to certain  
diseases or disease types. A further use for the invention is for  
developing methods useful in diagnosing disease.

Sequence 1210 BP: 403 A; 200 C; 234 G; 373 T; 0 other;

Query Match	10.6%	Score 107.2;	DB 16;	Length 1210;
Best Local Similarity	66.7%	Pred. No. 1.3e-14;		
Matches 184;	Conservative 0;	Mismatches 63;	Indels 29;	Gaps 1;

Qy	193	atatttccttatttttttaagaattattattattaaagtatatagtagaacactgcctctcc	252
Db	404	ATTTTCTATTTTTTTTAAAGAAATTTTTACTTTATGTGAGTAGACACCACCGCTGT	345
Qy	253	ctcagacatacagtagtacaggcatcgatccattaccagatggttgtagcaccaatgt	312
Db	344	CNTTGACACACAGAGAGGGCATCAGATCCCAITACAGATGTTGTGAGGCCACCATGT	285
Qy	313	ggtttcacagatggtttgagcccacatggtttcaggaattaaactcaggaccttgg	372
Db	284	GGTTCG-----TGGGAATTGAACTCAGGTCCTCTGG	254
Qy	373	asagacagtcagtctctaacccttaagccatctctcctgcccttatatacaaattta	432
Db	253	AAGACAGTCAGTGCTTTTAAC TGCTGAGGCATCTCTCCAGCTCTCATTTTCTATTTTA	194
Qy	433	atgctacgtacacacaactctcttctcccttaatgg	468
Db	193	GACCAATTAGGTTCCCATCCCACTTTCTACCTCTATGG	158

## RESULT

RESOLUT 4  
X55300  
ID X55300 standard: DNA: 48974 BP.

XX  
AC  
X55300:XX  
DT 19-JUL-1999 (first entry)XX  
DE Mouse Presentin-1 gene sequence.

XX Neuron-specific transcription; promoter; mouse; Presenilin-1 gene;  
KW transgenic; laboratory model; Alzheimer's disease; ss.  
KW

XX  
05  
Miss  
sp.XX  
PN  
W09910368-A1XX  
PD  
04-MAR-1999XX  
28-ATTG-1998-  
DFXX  
BB 28-APC-1997. 97MS-0920422

XX  
XX  
XX

[illegible]

XX 1000 3100 100

XX

XX

XX  
XX

transcription of a promoter segment in a mammalian cell. The promoter segment is from the mouse genomic Presenilin-1 gene. Transgenic non-human mammals containing a DNA expression cassette comprising the neuron-specific promoter are useful as laboratory models for studying the function of the Presenilin gene, and for studying the etiology of Alzheimer's disease.

Sequence 48974 BP: 12579 A: 10749 C: 11189 G: 14438 T: 19 other; XX

Query Match            10.4%; Score 105; DB 20; Length 48974;  
Best Local Similarity 66.1%; Pred. NO. 9.1e-14;  
Matches 183; Conservative 0; Mismatches 65; Indels 29;

QY 194 ttttctttattttttaagattttatttaatttaattgaatgatacactgcctctccc 253  
||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
pb 3532 ttttaaagaatttttaagggtttattttatttatatgatgaacatttgtagtctc 3591













FT	CDS	599..4306		
FT	/tag=	a		
FT	/product=	serine_ectopeptidase		
FT		849..850		
FT	/tag=	b		
FT	/label=	Alternative_splice_site		
FT	/note=	"A second clone was isolated which differed upstream of this site and was identical downstream of it (see #72653)"		
XX				
PX	W09635805-A2.			
PD	14-NOV-1996.			
XX				
PX	09-MAY-1996;	96WO-FR00700.		
PF				
PR	09-MAY-1995;	95FR-0005489.		
XX				
PA	(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.			
PX	Bambal RB, Bishop PB, Bourgeat P, Chan S, Ganellin CR;			
PI	Leblond B, Moore ANJ, Schwartz JC, Vargas F;			
PI	Lihua Z, Rose C;			
PX				
DR	WPI; 1996-518693/51.			
PX	P-PSDB; W21572.			
XX				
PT	Screening medicaments for treating disorders linked to inactivation of endogenous neuro-peptide(s) - by contacting candidate molecule with membrane tri:peptidyl-peptidase (homologue) and measuring enzyme activity			
CC	Disclosure: Pages 166-171; 212pp; French.			
XX	An enzyme with specificity for cholecystokinin (CCK) substrates (specifically the non-sulphated CCK8 and the CCK5 peptides) was purified from rat cerebral cortex membranes using high performance liquid chromatography. Based on amino acid sequence data from the purified enzyme, PCR primers were designed to amplify oligonucleotide probes A and B of lengths 350 and 380 nucleotides, respectively. Using these probes, two distinct clones were identified in a rat brain cDNA library. The first clone had the sequence given in T72653 and is the rodent homologue of human tripeptidylpeptidase II. The second clone had the present sequence which differs from the first clone in the 5' region, probably as a result of alternative splicing, and appears to encode a serine ectopeptidase. The rat enzymes are preferred for use in a new method of screening for medicaments for treating disorders linked to the inactivation or degradation (or being treatable by retarding physiological degradation) of endogenous neuropeptides. In the method, a candidate molecule is contacted with a membrane tripeptidylpeptidase or homologue, and enzyme activity is measured. Disorders linked to inactivation or degradation of endogenous neuropeptides, include food intake disorders, cognitive and motor disorders such as anorexia, schizophrenia, Parkinson's disease and depression, as well as gastrointestinal transit disorders such as irritable bowel syndrome, bulimia and pathological obesity.			
XX	Sequence 5109 BP; 1541 A; 969 C; 1061 G; 1538 T; 0 other;			
SQ				
Query Match            8.8%; Score 89; DB 17; Length 5109;				
Best Local Similarity   66.7%; Pred. No. 1.6e-10;				
Matches 158; Conservative     0; Mismatches 50; Indels     29; Gaps				
QY	194	tttttctattttttaagattattattattatttaagtatatgacgtacctgcctccc 253 		
Dd	326	ttttaattattttttttaagattattattattatttatgtacgtcagttagctgcc 385 		
QY	254	tccagacatagcagtcaggcgcatcgatccattacagatgggttgagccaccatgtg 313 		
Dd	386	ttcagacacaccaaagaaggcgcatcgattccattccattacagatgggttgagccaccatgtg 445 		
QY	314	gtttcacagatggttgtgagccacocatgtggttttcaggaattgaactcagagaccttggca 373		

[illegible]









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:10:09 ; Search time 141.08 Seconds  
(without alignments)  
1080.602 Million cell updates/sec

Title: US-09-101-423a-3  
Perfect score: 1008  
Sequence: 1 agagacagattcacatccc.....ttctgtgtctctcctgaac 1008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104.8	10.4	13011	4	US-08-791-849A-14
2	89	8.8	6727	5	US-08-629-643A-5
3	84.4	8.4	26700	2	US-08-472-217-1
4	84.4	8.4	26700	3	US-08-488-193-5
5	84.4	8.4	26700	5	US-08-760-534A-1
6	77.4	7.7	13146	4	US-08-724-354D-3
7	77.4	7.7	13146	5	US-09-270-984A-3
8	75.4	7.5	51259	5	US-08-781-891-209
9	75.4	7.5	51259	5	US-09-166-186-107
10	72.6	7.2	5687	3	US-08-380-403A-3
11	72.6	7.2	5687	3	US-08-895-628-3
12	72.6	7.2	5687	3	US-08-676-279-49
13	71.4	7.1	14507	5	US-08-785-150-1
14	69.8	6.9	7970	1	US-08-135-511-31
15	69.8	6.9	7997	2	US-08-187-453-31
16	69.8	6.9	7997	2	US-08-562-985A-7
17	69.2	6.9	12537	3	US-08-611-280-4
18	67	6.6	6727	5	US-08-629-643A-5
19	66.8	6.6	3293	1	US-07-923-976-1
20	66.8	6.6	6645	3	US-08-380-403A-4
21	66.8	6.6	6645	4	US-08-895-628-4
22	66.4	6.6	601	5	US-09-080-855-34
23	66.4	6.6	613	5	US-09-080-855-35
24	66.4	6.6	3364	4	US-08-735-609-9
25	66.4	6.6	3364	4	US-08-735-609-9
26	66.4	6.6	3364	5	US-09-315-372-9

C 27	66.4	6.6	3364	5	US-09-244-752-9	Sequence 9, Appli
C 28	66.4	6.6	3364	5	US-09-245-497-9	Sequence 9, Appli
C 29	65.2	6.5	3963	2	US-08-464-961-1	Sequence 1, Appli
C 30	65.2	6.5	3963	5	US-08-907-800A-1	Sequence 1, Appli
C 31	65.2	6.5	3963	6	PCT-US96-08233-1	Sequence 1, Appli
C 32	65	6.4	5159	4	US-08-146-930-3	Sequence 3, Appli
C 33	65	6.4	5159	6	PCT-US93-03993-3	Sequence 3, Appli
C 34	61.8	6.1	2498	1	US-07-943-843-3	Sequence 3, Appli
C 35	61.8	6.1	2498	2	US-08-347-003-3	Sequence 3, Appli
C 36	61.8	6.1	16442	5	US-08-781-891-208	Sequence 208, App
C 37	61.4	6.1	4765	5	US-08-936-135-21	Sequence 21, Appli
C 38	61.4	6.1	4780	5	US-08-936-135-23	Sequence 23, Appli
C 39	61.2	6.1	4533	5	US-08-863-790-27	Sequence 27, Appli
C 40	60.8	6.0	4698	1	US-07-807-043B-5	Sequence 5, Appli
C 41	60.8	6.0	4698	1	US-08-299-849B-5	Sequence 5, Appli
C 42	60.8	6.0	4898	4	US-08-142-368A-5	Sequence 5, Appli
C 43	60.8	6.0	4898	5	US-08-967-727-5	Sequence 5, Appli
C 44	60.6	6.0	535	2	US-08-686-878A-38	Sequence 38, Appli
C 45	60.2	6.0	10409	5	US-08-772-440-33	Sequence 33, Appli

## ALIGNMENTS

RESULT 1  
US-08-791-849A-14  
; Sequence 14, Application US/08791849A  
; Patent No. 5914449  
; GENERAL INFORMATION:  
; APPLICANT: Makoto MURASE et al.  
; TITLE OF INVENTION: Method for Increasing Storage  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08791,849A  
; FILING DATE: January 30, 1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX:  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 13011 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ANTI-SENSE: NC  
; ORIGINAL SOURCE:  
; ORGANISM: rat (Rattus norvegicus)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(3212..3218, 3766..3948, 5917..6008,















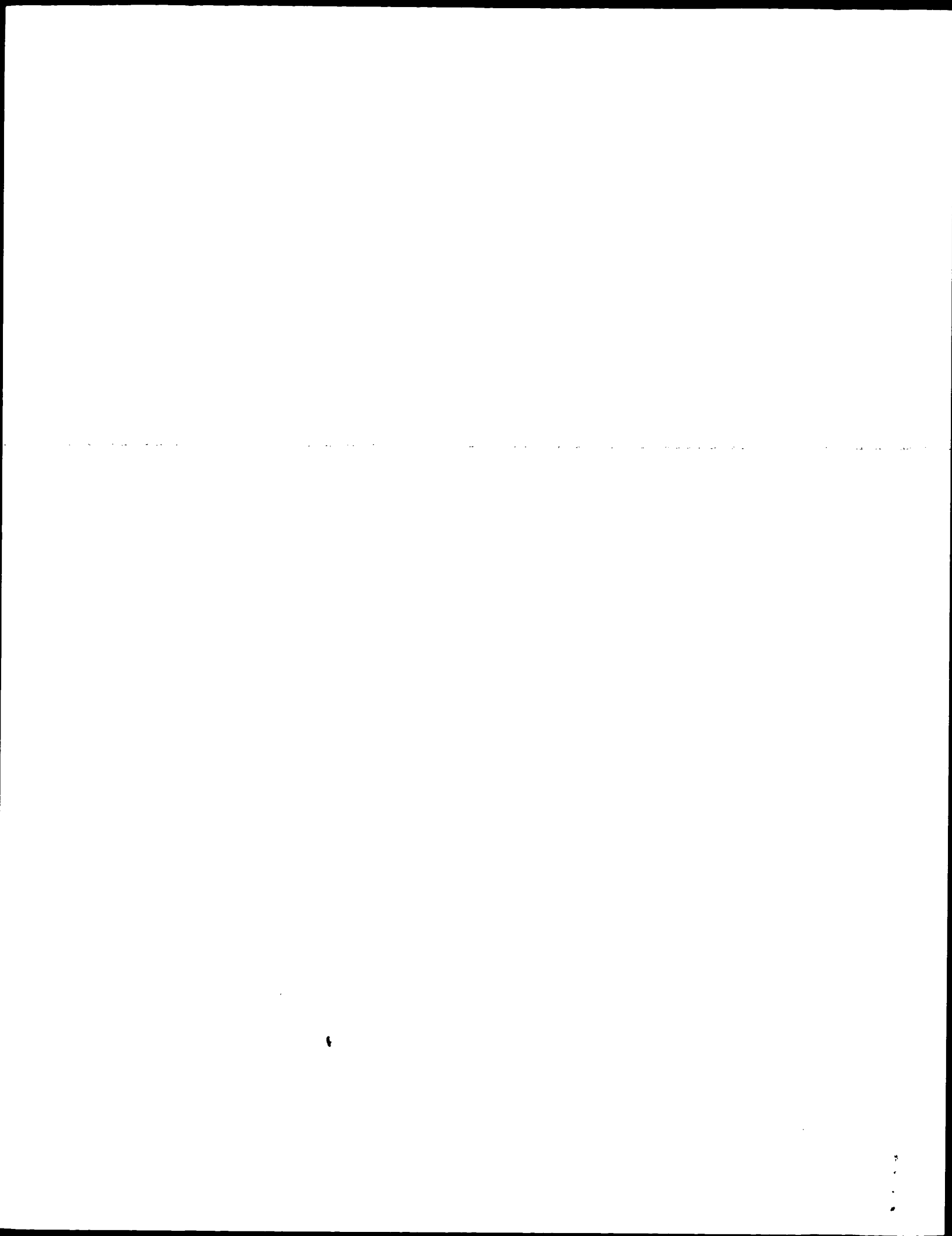


.. Mon Nov 6 10:14:22 2000

us-09-101-423a-3.rni

---

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 3, 2000, 22:07:46 ; Search time 3273.53 Seconds  
(without alignments)  
1903.837 Million cell updates/sec

Title: US-09-101-423A-3  
Perfect score: 1008  
Sequence: 1 agaccagattcacatccc.....ttccgtgtccctcctggaac 1008

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*  
44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estom:\*  
68: em\_estov1:\*  
69: em\_estov2:\*  
70: em\_estp11:\*  
71: em\_estp12:\*  
72: em\_estp13:\*  
73: em\_estp14:\*  
74: em\_estp15:\*  
75: em\_estro1:\*  
76: em\_estro2:\*  
77: em\_estro3:\*  
78: em\_estro4:\*  
79: em\_estro5:\*  
80: em\_estro6:\*  
81: em\_estro7:\*  
82: em\_estro8:\*  
83: em\_estro9:\*  
84: em\_estro10:\*  
85: em\_estro11:\*  
86: em\_estro12:\*  
87: em\_estro13:\*  
88: gb\_gss1:\*  
89: gb\_gss2:\*  
90: gb\_gss3:\*  
91: gb\_gss4:\*  
92: em\_gss1:\*  
93: em\_gss2:\*  
94: em\_gss3:\*  
95: em\_gss4:\*  
96: gb\_gss5:\*  
97: gb\_gss6:\*  
98: gb\_gss7:\*  
99: gb\_gss8:\*  
100: gb\_gss9:\*  
101: em\_gss5:\*  
102: em\_gss6:\*  
103: em\_gss7:\*  
104: em\_gss8:\*  
105: em\_gss9:\*  
106: em\_gss10:\*  
107: em\_gss11:\*  
108: gb\_gss10:\*  
109: gb\_gss11:\*  
110: em\_gss12:\*  
111: gb\_gss12:\*  
112: gb\_gss13:\*  
113: gb\_gss14:\*  
114: gb\_gss15:\*  
115: gb\_gss16:\*  
116: gb\_gss17:\*



Ddb	241	TTCCTCCTGACCCAGTGCATTAAGTAGAGACTTGTTCTATGACTTTGTAG	294
RESULT	2		
LOCUS	AA874895	342 bp	mrna EST 04-JUL-1999
DEFINITION	UI-R-E0-ck-e-03-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone		
AUTHORS	UI-R-E0-ck-e-03-0-UI 3' similar to gi 2291938 gb AA543461 AA543461		
TITLE	vj92b06.rl Soares mouse mammary gland NbMMG Mus musculus cDNA clone		
KEYWORDS	935507 5' similar to WP:R06F6.9 CEO1625 ENOYL COA		
ACCESSION	HYDRATASE/ISOMERASE ; , mRNA sequence.		
VERSION	AA874895		
SOURCE	AA874895.1 GI:4230777		
ORGANISM	Norway rat.		
REFERENCE	Rattus norvegicus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
JOURNAL	Rattus.		
MEDLINE	1 (bases 1 to 342)		
COMMENT	Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery Genome Res. 6 (9), 791-806 (1996) On Mar 20, 1998 this sequence version replaced gi:2979843. Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to identify it as a clone from the normalized adult 8-day-embryo library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone Distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE ID-177563 The following repetitive elements were found in this cDNA sequence: 1-219, >B2#SINE/B2 Seq primer: M13 Forward POLYA-No.		
FEATURES	source		
BASE COUNT	93 a 66 c 67 g 115 t 1 others		
ORIGIN			
Query Match	11.8%; Score 119; DB 7; Length 342;		
Best Local Similarity	74.3%; Pred. No. 1.3e-15;		
Matches 182; Conservative	0; Mismatches 50; Indels 13; Gaps 2;		
QY	194	ttttctttttttttaagattttatttttaagtattgtagtacactgcctctccc	253
Db	7	TTTTTTTGGTTTGAATACTTTTAAATATTATTATATGTTAGTAGTACACTGTAGCTGTC	66

















GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 04:10:19 ; Search time 7111.01 Seconds  
(without alignments)  
589.632 Million cell updates/sec

Title: US-09-101-423A-5  
Perfect score: 960  
Sequence: 1 gagggggtgggcacagtt.....tcaaccacacacacacat 960

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1033670 seqs, 2183789903 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pr1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: em\_fun.\*  
13: em\_hum1.\*  
14: em\_hum2.\*  
15: em\_in.\*  
16: em\_om.\*  
17: em\_or.\*  
18: em\_ov.\*  
19: em\_pat.\*  
20: em\_ph.\*  
21: em\_pl.\*  
22: em\_ro.\*  
23: em\_sts.\*  
24: em\_sy.\*  
25: em\_un.\*  
26: em\_vi.\*  
27: gb\_htg1.\*  
28: gb\_htg2.\*  
29: gb\_in1.\*  
30: gb\_in2.\*  
31: em\_ba1.\*  
32: em\_ba2.\*  
33: em\_hum3.\*  
34: em\_hum4.\*  
35: gb\_pr4.\*  
36: gb\_htg3.\*  
37: gb\_htg4.\*  
38: gb\_htg5.\*  
39: gb\_htg6.\*  
40: gb\_htg7.\*  
41: em\_htg1.\*  
42: em\_htg2.\*  
43: em\_htg3.\*

44: em\_hum5.\*  
45: gb\_pl3.\*  
46: gb\_pr5.\*  
47: gb\_htg8.\*  
48: gb\_htg9.\*  
49: gb\_htg10.\*  
50: gb\_htg11.\*  
51: gb\_htg12.\*  
52: gb\_htg13.\*  
53: gb\_htg14.\*  
54: gb\_in3.\*  
55: gb\_htg15.\*  
56: gb\_htg16.\*  
57: gb\_htg17.\*  
58: em\_htg4.\*  
59: em\_htg5.\*  
60: em\_htg6.\*  
61: em\_htg7.\*  
62: em\_hum6.\*  
63: gb\_htg18.\*  
64: gb\_htg19.\*  
65: gb\_ba3.\*  
66: em\_htg8.\*  
67: em\_htg9.\*  
68: em\_htg10.\*  
69: em\_htg11.\*  
70: em\_htg12.\*  
71: em\_htg13.\*  
72: em\_htg14.\*  
73: em\_htg15.\*  
74: em\_htg16.\*  
75: em\_htg17.\*  
76: em\_htg18.\*  
77: em\_htg19.\*  
78: em\_htg20.\*  
79: em\_htg21.\*  
80: em\_htg22.\*  
81: em\_htg23.\*  
82: gb\_pr6.\*  
83: gb\_pr7.\*  
84: gb\_htg20.\*  
85: gb\_htg21.\*  
86: gb\_htg22.\*  
87: gb\_htg23.\*  
88: gb\_ro.\*  
89: gb\_sts1.\*  
90: gb\_sts2.\*  
91: gb\_sy.\*  
92: gb\_un.\*  
93: gb\_vil.\*  
94: gb\_vil2.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	960	100.0	960	5	A64062	A64062 Sequence 5
2	36	3.8	175371	57	AC073748	AC073748 Mus muscu
3	36	3.8	233752	57	AC073683	AC073683 Mus muscu
4	35	3.6	112948	49	AC023806	AC023806 Mus muscu
5	28	2.9	2414	88	MMMS24	AJ002241 Mus muscu
6	28	2.9	215601	57	AC073806	AC073806 Mus muscu
7	28	2.9	225509	57	AC073782	AC073782 Mus muscu
8	27	2.8	207869	57	AC073697	AC073697 Mus muscu
9	26	2.7	1678	5	A51790	A51790 Sequence 25
10	26	2.7	13972	88	AF149204	AF149204 Mus muscu
11	26	2.7	21248	88	AF004854	AF004854 Mus muscu
12	26	2.7	110000	47	AC020850_3	Continuation (4 of

```
13 26 2.7 151340 47 AC020961 Mus muscu
14 26 2.7 162496 53 AC044846 Mus muscu
15 26 2.7 177491 52 AC034285 Mus muscu
16 26 2.7 196689 57 AC073730 Mus muscu
17 26 2.7 210650 50 AC025794 Mus muscu
18 26 2.7 219720 49 AC023364 Mus muscu
19 26 2.7 252504 57 AC073815 Mus muscu
20 25 2.6 93265 88 AP002736 Mus muscu
21 25 2.6 139171 50 AC024883 Mus muscu
22 25 2.6 157225 57 AC074359 Mus muscu
23 25 2.6 206126 38 AC015800 Mus muscu
24 25 2.6 208632 57 AC073711 Mus muscu
25 25 2.6 234787 57 AC073780 Mus muscu
26 25 2.6 249487 88 MMU251788 Mus muscu
27 25 2.6 249487 88 MMU251835 Mus muscu
28 24 2.5 168790 38 AC015886 Mus muscu
29 24 2.5 203201 38 AC015915 Mus muscu
30 24 2.5 219811 52 AC034123 Mus muscu
31 24 2.5 314146 57 AC073759 Mus muscu
32 24 2.5 314146 57 AC073759 Mus muscu
33 23 2.4 9754 88 MMU14422 Mus muscu
34 23 2.4 29802 27 AC007518 Mus muscu
35 23 2.4 139097 27 AC004591 Mus muscu
36 23 2.4 146612 56 AC073151 Mus muscu
37 23 2.4 176245 38 AC016017 Mus muscu
38 23 2.4 191672 47 AC020794 Mus muscu
39 23 2.4 194599 57 AC073668 Mus muscu
40 23 2.4 206051 50 AC024914 Mus muscu
41 23 2.4 219218 47 AC020817 Mus muscu
42 23 2.4 241679 47 AC020861 Mus muscu
43 22 2.3 94459 88 MMT5XDNA Mus muscu
44 22 2.3 109582 10 AC009482 Mus muscu
45 22 2.3 208632 57 AC073711 Mus muscu
```

## ALIGNMENTS

```
RESULT 1
LOCUS A64062 960 bp DNA
DEFINITION Sequence 5 from Patent WO9725443.
ACCESSION A64062
VERSION A64062.1 GI:3717536
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Rudland,P.S. and Barraclough,B.R.
TITLE METASTASIS INDUCING DNA'S
JOURNAL Patent: WO 9725443-A 17-JUL-1997;
UNIV LIVERPOOL (GB)
FEATURES
Source
1..960
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 256 a 193 c 233 g 278 t
ORIGIN
```

```
Query Match 100.0%; Score 960; DB 5; Length 960;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gaggggggtggggcacagtattgtttgtaggagggttccatgaacctcagcagct 60
Dy 1 GAGGGGTGGTGGGCACAGTATGTTTGTAGGAAGGGTCCAGAACCTTCACAGAGCT 60

Qy 61 cgggttagaatttaaagccctgaggggaatttttttaaatcgctatgaatcgac 120
Dy 61 CGGTTAGAAATTTAAAGCCCTGAGGGGAATTTTTTTTAAATCGCTATGAATCTGAC 120
```

```
Qy 121 atgagaaaaacagatcagaacgtttctgtgttcagaaaaaggacaaagtgtgtgctaa 180
Dy 121 ATGAGAAAACAGATCAGAAAGCTTCTGTGCTTCAGAAAAGGACAAAGTGTGTGAGCTAA 180

Qy 181 cagactgcacactgtgtgttcagggcacatcttgatcagagaggttcagataatgtcccc 240
Dy 181 CAGACTGCACACTGTTGTTCGAGGACACATCTGGATCAGAGGAGCGTCAGATAATGTCCCC 240

Qy 241 aaagttaaatgactgtgttcagacagtaacagagtggtgggggggtgctacagccacgc 300
Dy 241 AAAGTTAAATGACTTGTGTGCACAGTACCCAGTGTGTGGGGGTGCTACAGCCCCAGC 300

Qy 301 ggtctcaacctctctgatgttcagaccttttaacacagtgccctcatctctgttgacct 360
Dy 301 GGTCTCAACCTCTCTGATGCTTCGACCCCTTAAACAGTGCCTCATCTCTGTGTGACTT 360

Qy 361 ccccaaccttaaatatttttggctgttcataaactgtgatttttgatactgttatgaa 420
Dy 361 CCCCAACCTTAAATATTTTGTGCTGTTTCATAAATGATTTTGTATGATCTGTTATGAA 420

Qy 421 ttgtaataataataattttgaagaagaggtttgccaagggtttgagaactgctgttcta 480
Dy 421 TTGTAATAATAATAATTTGAGAAAGAGGTTTGCCAAAGGTTTGAGAACTGCTGTTCTA 480

Qy 481 gcccacagtggatggttttttgcgtcatttggggtttttatgaggcagagctcttatgacc 540
Dy 481 GCCCACAGTGGATGCTTTTCGTCATTTGGGGTTTTATGAGGAGAGCTCTTATGTAGCC 540

Qy 541 cagggtagcagcctagatgtgtctacttagctagtgaggataaacttggaactctgagac 600
Dy 541 CAGGCTAGCAGCCTAGATGTGCTACTTACTTAGCTGAGGAATAACCTTGGAACTTCTGAGGAC 600

Qy 601 tggagagactgctttagtccctcaagaactggaactggaactgagttgtgacttggtgg 660
Dy 601 TGGAGAGACTGCTTGTAGTCCCTCAGAAACTGGAAATAGCTGGAGTTTGGCTACTTGTGGG 660

Qy 661 ttcccttttctcacaaccttttctactcttttccacctgtcgccccccttaacactaaa 720
Dy 661 TTCCCTTTTCTCAAACTTTTCTACTCTTTTCCACCTGTGCGGCCCTTAACACTAAA 720

Qy 721 taagaagaagaaggggagcagatagaggggaaagaaacccctgaataacgtcagtagtg 780
Dy 721 TAAGAAGAGAAAGGGGAGCATAGAGGGGAAAGAAACCCCTGAATAACGTCAGTAGTTG 780

Qy 781 gcaaggggggtgacatatgtgtcattagaccacatctggtgattaaaggggagtaag 840
Dy 781 GCAAGGGGGGTGACATATGTTGTCATTAGACCACATCTCTGCTGATTAAGGGGAGTCAAG 840

Qy 841 ttccctgggggaagtttgatcttctgtgttaacgatataattttcttccctgttgcctt 900
Dy 841 TTCCCTGGGGGAAGTTTGTATCTTTCGTGTAACGATATCTAATTTCTTCTCCTGTTGCTT 900

Qy 901 cgtctttgaaacagcacttgataccccacaatggaccatcaaccacccaacccaacat 960
Dy 901 CGTCTTTGTGAACACGACTTGTATAACCCACATGGAGCCATCAACCAACCAACCAACAT 960
```

```
RESULT 2
LOCUS AC073748 175371 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-305J2, WORKING DRAFT SEQUENCE, 33 unordered
pieces.
ACCESSION AC073748
VERSION AC073748.1 GI:8810365
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 175371)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

2 (bases 1 to 175371)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 1854210  
Center clone name: RPCI-23\_30572  
-----

Summary Statistics  
Consensus quality: 152503 bases at least Q40  
Consensus quality: 164435 bases at least Q30  
Consensus quality: 166779 bases at least Q20  
Estimated insert size: 166550; agarose-fp estimation  
Quality coverage: 11.93 in Q20 bases; agarose-fp estimation  
Quality coverage: 11.54 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1183 1282: contig of 1182 bp in length  
1283 1282: gap of unknown length  
1283 2853: contig of 1571 bp in length  
2854 2953: gap of unknown length  
2954 4183: contig of 1230 bp in length  
4184 4283: gap of unknown length  
4284 5786: contig of 1503 bp in length  
5787 5886: gap of unknown length  
5887 7067: contig of 1180 bp in length  
7067 7167: gap of unknown length  
7167 8238: contig of 1072 bp in length  
8239 8338: gap of unknown length  
8339 9718: contig of 1380 bp in length  
9719 11510: contig of 1692 bp in length  
9819 11511: gap of unknown length  
11511 12959: contig of 1349 bp in length  
12960 13059: gap of unknown length  
13060 14954: contig of 1895 bp in length  
14955 15054: gap of unknown length  
15055 16822: contig of 1808 bp in length  
16863 16962: gap of unknown length  
16963 19187: contig of 2225 bp in length  
19188 19287: gap of unknown length  
19288 21631: contig of 2344 bp in length  
21632 21731: gap of unknown length  
21732 23582: contig of 1851 bp in length  
23583 23682: gap of unknown length  
23683 25787: contig of 2105 bp in length  
25788 25887: gap of unknown length  
25888 27450: contig of 1563 bp in length  
27451 27550: gap of unknown length  
27551 31646: contig of 4096 bp in length  
31647 33311: contig of 1565 bp in length  
33312 33411: gap of unknown length  
33412 36480: contig of 3069 bp in length  
36481 36581: gap of unknown length  
36581 40140: contig of 3560 bp in length  
40141 40240: gap of unknown length  
40241 46716: contig of 6476 bp in length  
46717 46816: gap of unknown length  
46817 52150: contig of 5334 bp in length

\* 52151 52250: gap of unknown length  
\* 52251 56774: contig of 4524 bp in length  
\* 56775 56874: gap of unknown length  
\* 56875 62585: contig of 5711 bp in length  
\* 62586 62685: gap of unknown length  
\* 62686 72379: contig of 9694 bp in length  
\* 72380 72479: gap of unknown length  
\* 72480 78605: contig of 6126 bp in length  
\* 78606 78705: gap of unknown length  
\* 78706 87900: contig of 9195 bp in length  
\* 87901 88000: gap of unknown length  
\* 88001 98832: contig of 10832 bp in length  
\* 98833 98932: gap of unknown length  
\* 98933 109561: contig of 10629 bp in length  
\* 109562 109661: gap of unknown length  
\* 109662 127036: contig of 17375 bp in length  
\* 127037 127136: gap of unknown length  
\* 127137 139467: contig of 12331 bp in length  
\* 139468 139567: gap of unknown length  
\* 139568 158423: contig of 18856 bp in length  
\* 158424 158523: gap of unknown length  
\* 158524 175371: contig of 16848 bp in length.

FEATURES  
source

1. 175371  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-30572"  
/clone\_lib="RPCI mouse BAC library 23"  
BASE COUNT 44046 a 42634 c 42369 g 43110 t 3212 others  
ORIGIN

Query Match 3.88; Score 36; DB 57; Length 175371;  
Best Local Similarity 100.0%; Pred. No. 5.5e-08;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 301 ggtctcaacctctgctgaccttcaata 336  
|||||  
Db 158716 GGTCTCAACCTTCTGATGCTGACCCCTTATA 158751

## RESULT 3

AC073683/3  
LOCUS AC073683.1 GI:8810300 HTG 29-JUN-2000  
DEFINITION Mus musculus clone RP23-124G7, WORKING DRAFT SEQUENCE, 23 unordered  
pieces.  
ACCESSION AC073683.1  
VERSION AC073683.1  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 233752)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Mouse  
JOURNAL Unpublished  
2 (bases 1 to 233752)  
REFERENCE DOE Joint Genome Institute.  
AUTHORS Direct Submission  
TITLE Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 1784639  
Center clone name: RPCI-23\_124G7  
-----  
Summary Statistics  
Consensus quality: 212896 bases at least Q40

Consensus quality: 224171 bases at least Q30  
 Consensus quality: 226565 bases at least Q20  
 Estimated insert size: 226000; agarose-fp estimation  
 Estimated insert size: 231552; sum-of-contigs estimation  
 Quality coverage: 7.43 in Q20 bases; agarose-fp estimation  
 Quality coverage: 7.25 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 23 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1159: contig of 1159 bp in length  
 1160 1250: gap of unknown length  
 1260 2669: contig of 1410 bp in length  
 2670 2769: gap of unknown length  
 2770 3792: contig of 1023 bp in length  
 3793 3892: gap of unknown length  
 3893 5278: contig of 1386 bp in length  
 5279 5378: gap of unknown length  
 5379 6486: contig of 1107 bp in length  
 6487 6586: gap of unknown length  
 6587 7907: contig of 1322 bp in length  
 7908 8008: gap of unknown length  
 8009 9687: contig of 1680 bp in length  
 9688 9787: gap of unknown length  
 9788 11694: contig of 1807 bp in length  
 11695 13423: contig of 1729 bp in length  
 13424 13524: gap of unknown length  
 13525 15543: contig of 2019 bp in length  
 15544 18366: gap of unknown length  
 18367 18467: contig of 2724 bp in length  
 18468 21545: gap of unknown length  
 21546 25552: contig of 3079 bp in length  
 25553 35623: contig of 3907 bp in length  
 35624 35723: gap of unknown length  
 35724 51559: contig of 9971 bp in length  
 51560 70197: gap of unknown length  
 70198 70297: contig of 15836 bp in length  
 70298 92176: contig of 18538 bp in length  
 92177 92276: contig of 21879 bp in length  
 92277 113150: gap of unknown length  
 113151 132250: contig of 20874 bp in length  
 132251 132933: gap of unknown length  
 132934 133033: contig of 19683 bp in length  
 133034 156488: gap of unknown length  
 156489 179518: contig of 23455 bp in length  
 179519 179518: gap of unknown length  
 179519 205622: contig of 22930 bp in length  
 205623 205722: contig of 26004 bp in length  
 205723 233752: gap of unknown length  
 233753 233752: contig of 28030 bp in length.

## FEATURES

Location/Qualifiers  
 1..233752  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-124C7"  
 /clone.lib="RP23 mouse BAC library 23"  
 BASE COUNT 59617 a 55573 c 57688 g 58669 t 2205 others  
 ORIGIN

Query Match 3.8%; Score 36; DB 57; Length 233752;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-08;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 agggagagcaagttcccttggggcaagtttgatcttt 864  
 Db 43089 AGGGGAGTCAAGTTCCTTGGGGCAAGTTTGATCTTT 43054

## RESULT 4

AC023806 LOCUS 112948 bp DNA HTG 15-JUN-2000  
 DEFINITION Mus musculus clone RP23-273C14, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 50  
 unordered pieces.  
 AC023806 ACCESSION  
 AC023806.6 GI:8248599  
 VERSION  
 HTG: HTG3\_PHASE1.  
 SOURCE  
 house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 112948)

## AUTHORS

Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,  
 Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,  
 Addison, S., Face, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,  
 Buha, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,  
 Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,  
 Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,  
 Gorrell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,  
 Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,  
 Kovar, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Martin, R.,  
 Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,  
 Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,  
 Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,  
 Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,  
 Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,  
 Worley, K. and Gibbs, R.

## TITLE

Direct Submission

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 112948)

## AUTHORS

Worley, K.C.

## TITLE

Direct Submission

## JOURNAL

Submitted (18-FEB-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Jun 5, 2000 this sequence version replaced gi:7158027.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: MAM

Center clone name: RP23-273C14

----- Summary Statistics

Sequencing vector: M13: L08821

Chemistry: Dye-terminator Big Dye: 24% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 66849 bases at least Q40

Consensus quality: 93367 bases at least Q30

Consensus quality: 104792 bases at least Q20

Estimated insert size: 92385; sum-of-contigs estimation

Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 50 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 7387: contig of 7387 bp in length

\* 7388 7487: gap of unknown length

\* 7488 12778: contig of 5291 bp in length

\* 12779 12878: gap of unknown length



```

* 12879 18944: contig of 6066 bp in length
* 18945 19044: gap of unknown length
* 19045 22462: contig of 4418 bp in length
* 22462 23562: gap of unknown length
* 23562 27102: contig of 3540 bp in length
* 27102 30927: contig of 3725 bp in length
* 30927 31027: gap of unknown length
* 31027 34969: contig of 3942 bp in length
* 34969 35069: gap of unknown length
* 35069 38711: contig of 3642 bp in length
* 38711 38811: gap of unknown length
* 38811 41745: contig of 2934 bp in length
* 41745 41845: gap of unknown length
* 41845 44231: contig of 2386 bp in length
* 44231 44331: gap of unknown length
* 44331 48279: contig of 3948 bp in length
* 48279 48379: gap of unknown length
* 48379 50919: contig of 2540 bp in length
* 50919 51019: gap of unknown length
* 51019 53247: contig of 2227 bp in length
* 53247 53347: gap of unknown length
* 53347 55660: contig of 2313 bp in length
* 55660 57773: gap of unknown length
* 57773 57873: contig of 2014 bp in length
* 57873 59529: contig of 1656 bp in length
* 59529 61846: gap of unknown length
* 61846 61946: contig of 2217 bp in length
* 61946 64542: gap of unknown length
* 64542 64642: gap of unknown length
* 64642 66044: contig of 1402 bp in length
* 66044 66144: gap of unknown length
* 66144 67853: contig of 1709 bp in length
* 67853 70002: contig of 2049 bp in length
* 70002 70102: gap of unknown length
* 70102 72286: contig of 2184 bp in length
* 72286 72386: gap of unknown length
* 72386 74293: contig of 1907 bp in length
* 74293 74393: gap of unknown length
* 74393 75757: contig of 1364 bp in length
* 75757 75857: gap of unknown length
* 75857 77244: contig of 1387 bp in length
* 77244 77344: gap of unknown length
* 77344 78608: contig of 1264 bp in length
* 78608 80143: gap of unknown length
* 80143 80243: contig of 1435 bp in length
* 80243 81846: gap of unknown length
* 81846 81946: contig of 1603 bp in length
* 81946 83606: gap of unknown length
* 83606 85147: contig of 1441 bp in length
* 85147 85247: gap of unknown length
* 85247 86756: contig of 1509 bp in length
* 86756 86856: gap of unknown length
* 86856 87882: contig of 1026 bp in length
* 87882 89554: gap of unknown length
* 89554 89654: contig of 1572 bp in length
* 89654 90855: gap of unknown length
* 90855 90955: contig of 1201 bp in length
* 90955 92285: gap of unknown length
* 92285 92386: contig of 1330 bp in length
* 92386 93764: gap of unknown length
* 93764 93765: contig of 1279 bp in length
* 93765 95222: contig of 1458 bp in length
* 95222 95322: gap of unknown length
* 95322 96707: contig of 1385 bp in length
* 96707 96808: gap of unknown length
* 96808 98324: contig of 1517 bp in length

```

```

* 98325 98424: gap of unknown length
* 98425 99727: contig of 1303 bp in length
* 99727 100924: gap of unknown length
* 100924 101024: contig of 1097 bp in length
* 101024 102177: gap of unknown length
* 102177 102277: contig of 1153 bp in length
* 102277 103278: gap of unknown length
* 103278 103339: contig of 1262 bp in length
* 103339 103540: gap of unknown length
* 103540 104872: contig of 1233 bp in length
* 104872 104972: gap of unknown length
* 104972 106007: contig of 1035 bp in length
* 106007 106107: gap of unknown length
* 106107 107203: contig of 1096 bp in length
* 107203 107303: gap of unknown length
* 107303 108512: contig of 1209 bp in length
* 108512 108612: gap of unknown length
* 108612 110018: contig of 1406 bp in length
* 110018 110118: gap of unknown length
* 110118 111535: contig of 1416 bp in length
* 111535 111634: gap of unknown length
* 111634 112948: contig of 1314 bp in length.

```

# FEATURES

## source

```

1. .112948
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-273C14"

```

BASE COUNT 26153 a 27405 c 27695 g 26758 t 4937 others  
ORIGIN

Query Match 3.6%; Score 35; DB 49; Length 112948;  
Best Local Similarity 100.0%; Pred. No. 1.9e-07;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 ttctaaccttctgatgcttcgaccttataac 337  
|||||  
Db 53762 TTCTCAACCTCTCTGATGCTCGACCTTTAATAC 53796

## RESULT 5

MMMS24/c 2414 bp DNA ROD 27-JUL-1998  
LOCUS Mus musculus minisatellite tandem repeat (MMS24).  
DEFINITION  
ACCESSION AJ002241  
VERSION AJ002241.1 GI:3355784  
KEYWORDS  
SOURCE  
ORGANISM

House mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 2414)  
Bois P., Stead J.D., Bakshi S., Williamson J., Neumann R.,  
Moghadassadeh B. and Jeffreys A.J.  
Isolation and characterization of mouse minisatellites  
Genomics 50 (3), 317-330 (1998)

## TITLE

Journal  
MEDLINE  
98341115

## FEATURES

source  
1. .2414  
/organism="Mus musculus"  
/strain="Balb/c"  
/db\_xref="taxon:10090"  
/chromosome="7"  
/map="22cm"

## repeat\_region

1. .2414  
/note="minisatellite"  
/rpt\_type="RANDOM"

BASE COUNT 481 a 584 c 677 g 588 t 84 others

## ORIGIN

Query Match 2.9%; Score 28; DB 88; Length 2414;

Best Local Similarity 100.0%; Pred. No. 0.001; Mismatches 0; Indels 0; Gaps 0;

Qy 836 tcaagttcttgggcaagttgatctt 863

Db 572 tcaagttcttgggcaagttgatctt 545

## RESULT 6

AC073806 215601 bp DNA HTG 18-JUL-2000  
Mus musculus clone RP23-47N6, WORKING DRAFT SEQUENCE, 22 ordered pieces.

## AC073806

VERSION AC073806.2 GI:9256808  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 215601)

DOE Joint Genome Institute.

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 215601)

DOE Joint Genome Institute.

## JOURNAL

Direct Submission

## COMMENT

Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Jul 18, 2000 this sequence version replaced gi:8810423.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>

## Project Information

Center Project Name: 1755238

Center clone name: RPCI-23\_47N6

## Summary Statistics

Consensus quality: 203615 bases at least Q40

Consensus quality: 211090 bases at least Q30

Consensus quality: 212651 bases at least Q20

Estimated insert size: 220390; agarose-fp estimation

Estimated insert size: 214601; sum-of-contigs estimation

Quality coverage: 7.88 in Q20 bases; agarose-fp estimation

Quality coverage: 8.1 in Q20 bases; sum-of-contigs estimation.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 22 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 4864: contig of 4864 bp in length

\* 4865 4984: gap of unknown length

\* 4965 21161: contig of 16197 bp in length

\* 21162 21261: gap of unknown length

\* 21262 22073: contig of 812 bp in length

\* 22074 22173: gap of unknown length

\* 22174 33484: contig of 11311 bp in length

\* 33485 33584: gap of unknown length

\* 33585 39528: contig of 5944 bp in length

\* 39529 39628: gap of unknown length

\* 39629 62157: contig of 22529 bp in length

\* 62158 62257: gap of unknown length

\* 62258 76745: contig of 14488 bp in length

\* 76746 76845: gap of unknown length  
\* 76846 81102: contig of 4257 bp in length  
\* 81103 81202: gap of unknown length  
\* 81203 85706: contig of 4504 bp in length  
\* 85707 85806: gap of unknown length  
\* 85807 90755: contig of 4949 bp in length  
\* 90756 90855: gap of unknown length  
\* 90856 101543: contig of 10688 bp in length  
\* 101544 101643: gap of unknown length  
\* 101644 119328: contig of 17685 bp in length  
\* 119329 119428: gap of unknown length  
\* 119429 122689: contig of 3261 bp in length  
\* 122690 122789: gap of unknown length  
\* 122790 131818: contig of 9029 bp in length  
\* 131819 131918: gap of unknown length  
\* 131919 150107: contig of 18189 bp in length  
\* 150108 150207: gap of unknown length  
\* 150208 156868: contig of 6661 bp in length  
\* 156869 156968: gap of unknown length  
\* 156969 187723: contig of 30755 bp in length  
\* 187724 187823: gap of unknown length  
\* 187824 190416: contig of 2593 bp in length  
\* 190417 190516: gap of unknown length  
\* 190517 195936: contig of 5420 bp in length  
\* 195937 196036: gap of unknown length  
\* 196037 204418: contig of 8382 bp in length  
\* 204419 204518: gap of unknown length  
\* 204519 213898: contig of 9380 bp in length  
\* 213899 213998: gap of unknown length  
\* 213999 215601: contig of 1603 bp in length.

## FEATURES

## source

1. 215601  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-47N6"  
/clone\_lib="RPCI mouse BAC library 23"

BASE COUNT 52000 a 52880 c 58600 g 50020 t 2101 others  
ORIGIN

## Query Match

Best Local Similarity 2.9%; Score 28; DB 57; Length 215601;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 836 tcaagttcttgggcaagttgatctt 863

Db 37362 tcaagttcttgggcaagttgatctt 37389

## RESULT 7

AC073782/c

## LOCUS

DEFINITION Mus musculus clone RP23-399G18, WORKING DRAFT SEQUENCE, 10 ordered pieces.

AC073782

VERSION AC073782.2 GI:9256794

KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Jul 13, 2000 this sequence version replaced gi:8810399.

-----Genome Center

Center: Joint Genome Institute

Center: Joint Genome Institute

Center: Joint Genome Institute

Center: Joint Genome Institute

Center: Joint Genome Institute

Center: Joint Genome Institute

Center: Joint Genome Institute

```

Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1890250
Center clone name: RPCI-23_399G18
-----
Summary Statistics
Consensus quality: 218629 bases at least Q40
Consensus quality: 223533 bases at least Q30
Consensus quality: 224265 bases at least Q20
Estimated insert size: 225109; agarose-fp estimation
Estimated insert size: 225109; sum-of-contigs estimation
Quality coverage: 8.0 in Q20 bases; agarose-fp estimation
Quality coverage: 8.58 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 7777: contig of 7777 bp in length
7778 7877: gap of unknown length
7878 15718: contig of 7841 bp in length
15719 15818: gap of unknown length
15819 22934: contig of 7116 bp in length
22935 23035: gap of unknown length
23036 88704: contig of 65670 bp in length
88705 133868: gap of unknown length
133869 133969: gap of unknown length
133970 135572: contig of 1604 bp in length
135573 135672: gap of unknown length
135673 179417: contig of 43745 bp in length
179418 179517: gap of unknown length
179518 199978: contig of 20461 bp in length
199979 200078: gap of unknown length
200079 222948: contig of 22870 bp in length
222949 223048: gap of unknown length
223049 225509: contig of 2461 bp in length.
*
* Location/Qualifiers
1. 225509
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-399G18"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 56682 a 56806 c 54655 g 56465 t 901 others
ORIGIN

Query Match 2.9%; Score 28; DB 57; Length 225509;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 tcaagtctcttgggcaagttgattt 863
|||||
Db 43008 TCAAGTCTCTTGGGCAAGTTGATCTT 42981
|||||

RESULT 8
AC073697 AC073697 207869 bp DNA HTG 29-JUN-2000
LOCUS Mus musculus clone RP23-161111, WORKING DRAFT SEQUENCE, 30
DEFINITION unorderd pieces.
ACCESSION AC073697
VERSION AC073697.1 GI:8810314
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 207869)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 207869)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1798899
Center clone name: RPCI-23_161111
-----
Summary Statistics
Consensus quality: 185896 bases at least Q40
Consensus quality: 197156 bases at least Q30
Consensus quality: 199406 bases at least Q20
Estimated insert size: 194000; agarose-fp estimation
Estimated insert size: 204969; sum-of-contigs estimation
Quality coverage: 8.57 in Q20 bases; agarose-fp estimation
Quality coverage: 8.11 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1615: contig of 1615 bp in length
1616 1715: gap of unknown length
1716 3062: contig of 1347 bp in length
3063 3163: gap of unknown length
3163 4489: contig of 1327 bp in length
4490 4590: gap of unknown length
4590 6764: contig of 2174 bp in length
6764 6864: gap of unknown length
6864 7949: contig of 1085 bp in length
7949 8048: gap of unknown length
8048 12380: contig of 4332 bp in length
12381 12480: gap of unknown length
12481 15171: contig of 2691 bp in length
15172 15271: gap of unknown length
15272 17274: contig of 2003 bp in length
17275 17375: gap of unknown length
17375 20360: contig of 2986 bp in length
20361 20461: gap of unknown length
20461 23055: contig of 2595 bp in length
23056 23156: gap of unknown length
23156 25147: contig of 1991 bp in length
25147 25247: gap of unknown length
25247 29518: contig of 4272 bp in length
29519 29619: gap of unknown length
29619 33008: contig of 3390 bp in length
33009 33108: gap of unknown length
33109 36438: contig of 3330 bp in length
36439 36538: gap of unknown length
36539 42339: contig of 5801 bp in length
42340 46521: contig of 4082 bp in length
46522 51088: contig of 4467 bp in length
51089 51189: gap of unknown length
51189 59971: contig of 8783 bp in length
59972 66942: contig of 6871 bp in length
66943 67042: gap of unknown length

```

```

* 67043 71805: contig of 4763 bp in length
* 71806 71905: gap of unknown length
* 71906 77944: contig of 6039 bp in length
* 77945 78044: gap of unknown length
* 78045 85953: contig of 7909 bp in length
* 85954 86054: gap of unknown length
* 86054 96320: contig of 10267 bp in length
* 96321 96420: gap of unknown length
* 96421 102614: contig of 6194 bp in length
* 102615 102714: gap of unknown length
* 102715 108050: contig of 5336 bp in length
* 108051 108150: gap of unknown length
* 108151 115865: contig of 7715 bp in length
* 115866 115965: gap of unknown length
* 115966 125992: contig of 10027 bp in length
* 125993 126092: gap of unknown length
* 126093 139568: contig of 13476 bp in length
* 139569 139668: gap of unknown length
* 139669 155125: contig of 15457 bp in length
* 155126 155225: gap of unknown length
* 155226 207869: contig of 52644 bp in length.
FEATURES
    source
    Location/Qualifiers
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone_lib="RP23-16111"
        /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 52448 a 48520 c 48115 g 55879 t 2907 Others
ORIGIN

```

```

Query Match      2.8%; Score 27; DB 57; Length 207869;
Best Local Similarity 100.0%; Pred. No. 0.0037;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 836 tcaagtctctgggcaagtttgatct 862
|||||
DB 97159 TCAGTCTCTGGGCAAGTTTGATCT 97185

```

```

RESULT 9
A51790 1678 bp DNA PAT 10-MAR-1997
LOCUS
DEFINITION Sequence 25 from Patent WO9617865.
ACCESSION A51790
VERSION A51790.1 GI:2304552
KEYWORDS
SOURCE
ORGANISM
REFERENCE
    1 (bases 1 to 1678)
    Mueller, H., Gillen, C. and Gleichmann, M.
    GENES AND GENE PRODUCTS FOR USE IN THE DIAGNOSIS OF DEGENERATIVE
    NERVE DAMAGE
    Patent: WO 9617865-A 25 13-JUN-1996;
    BOEHRINGER MANNHEIM GMBH (DE)
    Other publication AU 4302396 960626
    Other publication DE 19502525 960613.
FEATURES
    source
    Location/Qualifiers
        /organism="unidentified"
        /db_xref="taxon:32644"
BASE COUNT 427 a 380 c 443 g 427 t 1 others
ORIGIN

```

```

Query Match      2.7%; Score 26; DB 5; Length 1678;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 839 agtctctgggcaagtttgatcttt 864
|||||
DB 1229 AGTCTCTGGGCAAGTTTGATCTTT 1254

```

```

RESULT 10
AF149204/c
LOCUS
DEFINITION Mus musculus Su(var)3-9 homolog Suv39h2 (Suv39h2) gene, partial
cgs.
ACCESSION AF149204
VERSION AF149204.1 GI:8131951
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1 (bases 1 to 13972)
    O'Carroll, D., Lebersorger, A., and Jenuwein, T.
    Suv39h2 maps to chromosome 2
    Unpublished
REFERENCE
    2 (bases 1 to 13972)
    Laible, G., Lebersorger, A., O'Carroll, D. and Jenuwein, T.
    Direct Submission
    Submitted (07-MAY-1999) IMP, Dr. Bohrgasse 7, Vienna A1030, Austria
FEATURES
    Location/Qualifiers
        1..13972
        /organism="Mus musculus"
        /strain="129/Sv"
        /db_xref="taxon:10090"
        /chromosome="2"
        /map="2.9+/-0.4cm from centromere"
        join(<1437..1455,4003..4149,11746..>12402)
        /gene="Suv39h2"
        /product="Su(var)3-9 homolog Suv39h2"
        /protein_id="AA073152.1"
        /db_xref="GI:8131952"
        /translation="MASDLRTWCVPCLVSLDTLOELCRKEKLTCKSIGITKRNLLNVE
        KEVLCDDYKVAKEVYLVKMGKPDSTNTWEPLRLNRCPLLROFSDDKKTYLAGERKC
        YAVNSKSLQPAIAEVIYQAKQRIALQWQDYLNRRKNHKGMIENVVDLEGPPLOF
        YXINEYRPAGISINSEATSCCTDCFFDKCCPAEAGVVIAYNKKQOIKIQGTPIY
        ECNSRCRGPECPNRIQVKGQYSLCIFKTNNCGGKGVTLVKIKRMSFVMEYVGE"
BASE COUNT 3799 a 2789 c 3011 g 4308 t 65 others
ORIGIN

```

```

Query Match      2.7%; Score 26; DB 88; Length 13972;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 409 tactgttatgaattgtaataataa 434
|||||
DB 13459 TACTGTATGAATTGTAATAATAA 13434

```

```

RESULT 11
AF004854
LOCUS
DEFINITION Mus musculus Mmr52 gene, complete cds.
ACCESSION AF004854
VERSION AF004854.1 GI:2352489
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1 (bases 1 to 21248)
    Sunjevaric, I., Bendixen, C., Mortensen, U.H., Miljkovic, V. and
    Rothstein, R.

```



ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 151340)  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RG-MEAC\_247N19"  
 BASE COUNT 42206 a 33290 c 33063 g 42696 t 85 others  
 ORIGIN  
 Query Match 2.7% Score 26; DB 47; Length 151340;  
 Best Local Similarity 100.0%; Pred. No. 0.013;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 409 tactgttatgaattgtaataataa 434  
 Db 45557 TACTGTTATGAATTGTAATAATAA 45582  
 RESULT 14  
 AC044846/c  
 LOCUS  
 DEFINITION Mus musculus chromosome 11 clone RP23-46J7 map 11, WORKING DRAFT  
 AC044846  
 ACCESSION  
 VERSION  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 162496)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Mus musculus chromosome 11, clone RP23-46J7  
 Unpublished  
 Birren, B. 1 (bases 1 to 162496)  
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
 Boguslavsky, L., Boukhgaiter, B., Brown, A., Burkett, G.,  
 Campopiano, A., Castile, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, J.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Olivar, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 22, 2000 this sequence version replaced gi:7543812.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://f.p.genome.washington.edu/RM/RepeatMasker.html  
 TITLE  
 JOURNAL  
 COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 Project Information  
 Center project name: L9174  
 Center clone name: 46\_J7  
 Summary Statistics

Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 149639 bases at least Q40  
Consensus quality: 156386 bases at least Q30  
Consensus quality: 158956 bases at least Q20  
Insert size: 164000; agarose-fp  
Insert size: 160296; sum-of-contigs  
Quality coverage: 3.5 in Q20 bases; agarose-fp  
Quality coverage: 3.6 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1373: contig of 1373 bp in length  
1374 1473: gap of 100 bp  
1474 2786: contig of 1313 bp in length  
2787 2886: gap of 100 bp  
2887 4262: contig of 1376 bp in length  
4263 4362: gap of 100 bp  
4363 6417: contig of 2055 bp in length  
6418 6517: gap of 100 bp  
6518 9412: contig of 2895 bp in length  
9413 9512: gap of 100 bp  
9513 10790: contig of 1278 bp in length  
10791 10890: gap of 100 bp  
10891 15080: contig of 4190 bp in length  
15081 15180: gap of 100 bp  
15181 18824: contig of 3644 bp in length  
18825 18924: gap of 100 bp  
18925 23129: contig of 4205 bp in length  
23130 23229: gap of 100 bp  
23230 26859: contig of 3630 bp in length  
26860 26959: gap of 100 bp  
26960 32575: contig of 5616 bp in length  
32576 32675: gap of 100 bp  
32676 38781: contig of 6106 bp in length  
38782 38881: gap of 100 bp  
38882 44764: contig of 5883 bp in length  
44765 44864: gap of 100 bp  
44865 52115: contig of 7251 bp in length  
52116 52215: gap of 100 bp  
52216 59927: contig of 7712 bp in length  
59928 60027: gap of 100 bp  
60028 67115: contig of 7088 bp in length  
67116 67215: gap of 100 bp  
67216 75188: contig of 7973 bp in length  
75189 75288: gap of 100 bp  
75289 85641: contig of 10353 bp in length  
85642 85741: gap of 100 bp  
85742 97849: contig of 12108 bp in length  
97850 97949: gap of 100 bp  
97950 109572: contig of 11623 bp in length  
109573 109672: gap of 100 bp  
109673 122441: contig of 12769 bp in length  
122442 122541: gap of 100 bp  
122542 140243: contig of 17702 bp in length  
140244 140343: gap of 100 bp  
140344 162496: contig of 22153 bp in length.  
Location/Qualifiers  
1. .162496  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/map="11"  
/clone="RP23-46J7"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
1. .1373

FEATURES  
source

misc\_feature

misc\_feature /note="assembly\_fragment"  
1474. .2786  
/note="assembly\_fragment"  
2887. .4262  
/note="assembly\_fragment"  
4363. .6417  
/note="assembly\_fragment"  
6518. .9412  
/note="assembly\_fragment"  
9513. .10790  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
10891. .15080  
/note="assembly\_fragment"  
15181. .18824  
/note="assembly\_fragment"  
18925. .23129  
/note="assembly\_fragment"  
23230. .26859  
/note="assembly\_fragment"  
26960. .32575  
/note="assembly\_fragment"  
32676. .38781  
/note="assembly\_fragment"  
38882. .44764  
/note="assembly\_fragment"  
44865. .52115  
/note="assembly\_fragment"  
52216. .59927  
/note="assembly\_fragment"  
60028. .67115  
/note="assembly\_fragment"  
67216. .75188  
/note="assembly\_fragment"  
75289. .85641  
/note="assembly\_fragment"  
85742. .97849  
/note="assembly\_fragment"  
97950. .109572  
/note="assembly\_fragment"  
109673. .122441  
/note="assembly\_fragment"  
122542. .140243  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:left  
140344. .162496  
/note="assembly\_fragment"  
BASE COUNT 42470 a 37212 c 36404 g 44204 t 2206 others  
ORIGIN

Query Match 2.7% Score 26; DB 53; Length 162496;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 409 tactgttatgaattgaataataa 434  
|||||  
DB 38302 TACGTTATGATGTTATATAATA 38277

RESULT 15  
AC034285/c AC034285 177491 bp DNA HTG 22-JUL-2000  
LOCUS Mus musculus chromosome 11 clone RP23-220G21 map 11, WORKING DRAFT  
DEFINITION SEQUENCE, 30 unordered pieces.  
ACCESSION AC034285  
VERSION AC034285.2 GI:9369527  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 177491)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Mus musculus chromosome 11, clone RP23-220G21  
Unpublished  
2 (bases 1 to 177491)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,  
Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G.,  
Campomiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., DeAtrellano, K., Dewar, K., Diaz, J. S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
McLind, J., Meneus, L., Mihova, T., Miranda, C., Meng, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Olivier, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Trefry, S., Theodore, J., Tirrelli, A., Travers, M., Triglio, J.,  
Wassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 22, 2000 this sequence version released  
All repeats were identified using RepeatMasker:  
Smit, A. F. A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1965, Vol. 68, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Technology	1970, Vol. 1, No. 2, pp. 1-10
3. The Importance of Parental Involvement	Journal of Educational Psychology	1975, Vol. 66, No. 3, pp. 1-10
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 4, pp. 1-10
5. The Role of the School in the Community	Journal of Educational Research	1985, Vol. 88, No. 5, pp. 1-10
6. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	1990, Vol. 93, No. 6, pp. 1-10
7. The Importance of Teacher Professionalism	Journal of Educational Research	1995, Vol. 98, No. 7, pp. 1-10
8. The Effect of Teacher Salary on Student Achievement	Journal of Educational Research	2000, Vol. 103, No. 8, pp. 1-10
9. The Role of the Teacher in the 21st Century	Journal of Educational Research	2005, Vol. 108, No. 9, pp. 1-10
10. The Impact of Teacher Training on Student Achievement	Journal of Educational Research	2010, Vol. 113, No. 10, pp. 1-10

*	9018	9147:	gap of	100 bp
*	9118	11270:	contig of 2123 bp	in length
*	11271	11370:	gap of	100 bp
*	11371	13703:	contig of 2333 bp	in length
*	13704	13803:	gap of	100 bp
*	13824	15871:	contig of 3068 bp	in length
*	16872	16971:	gap of	100 bp
*	16972	19133:	contig of 2162 bp	in length
*	19134	19233:	gap of	100 bp
*	19234	22231:	contig of 2998 bp	in length
*	22232	22331:	gap of	100 bp
*	22332	25181:	contig of 3850 bp	in length
*	26132	26281:	gap of	100 bp
*	26232	29449:	contig of 3168 bp	in length
*	29430	29549:	gap of	100 bp
*	29530	33856:	contig of 4307 bp	in length
*	33857	33956:	gap of	100 bp
*	33957	37947:	contig of 3991 bp	in length
*	37918	38047:	gap of	100 bp
*	38018	41868:	contig of 3821 bp	in length
*	41869	41968:	gap of	100 bp
*	41939	46321:	contig of 4353 bp	in length
*	46322	46421:	gap of	100 bp
*	46432	50303:	contig of 3882 bp	in length
*	50334	50403:	gap of	100 bp
*	50434	57011:	contig of 6608 bp	in length
*	57012	57111:	gap of	100 bp
*	57112	62745:	contig of 5634 bp	in length
*	62716	62845:	gap of	100 bp
*	62846	69054:	contig of 6209 bp	in length
*	69055	69154:	gap of	100 bp
*	69155	75147:	contig of 5993 bp	in length
*	75148	75247:	gap of	100 bp
*	75248	82842:	contig of 7595 bp	in length
*	82843	82942:	gap of	100 bp
*	82913	90759:	contig of 7817 bp	in length
*	90750	90859:	gap of	100 bp
*	90850	102877:	contig of 12018 bp	in length
*	102878	102977:	gap of	100 bp
*	102978	115507:	contig of 12530 bp	in length
*	115508	115607:	gap of	100 bp
*	115608	128821:	contig of 13214 bp	in length
*	128822	128921:	gap of	100 bp
*	128922	138330:	contig of 9409 bp	in length
*	138331	138430:	gap of	100 bp
*	138431	153959:	contig of 15529 bp	in length
*	153950	154059:	gap of	100 bp
*	154050	177491:	contig of 23432 bp	in length.
FEATURES	Location/Qualifiers			
source	1. .177491			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/chromosome="11"			
	/map="11"			
	/clone="RP23-220G21"			
	/clone_lib="RPC1-23 Female Mouse BAC"			
misc_feature	1. .1136			
	/note="assembly_fragment"			
misc_feature	1237. .2279			
	/note="assembly_fragment"			
misc_feature	2380. .4178			
	/note="assembly_fragment"			
misc_feature	4279. .6004			
	/note="assembly_fragment"			
misc_feature	6105. .7395			
	/note="assembly_fragment"			
misc_feature	7496. .9047			
	/note="assembly_fragment"			
misc_feature	9148. .11270			
	/note="assembly_fragment"			
misc_feature	11371. .13703			
	/note="assembly_fragment"			
misc_feature	13804. .16871			



```

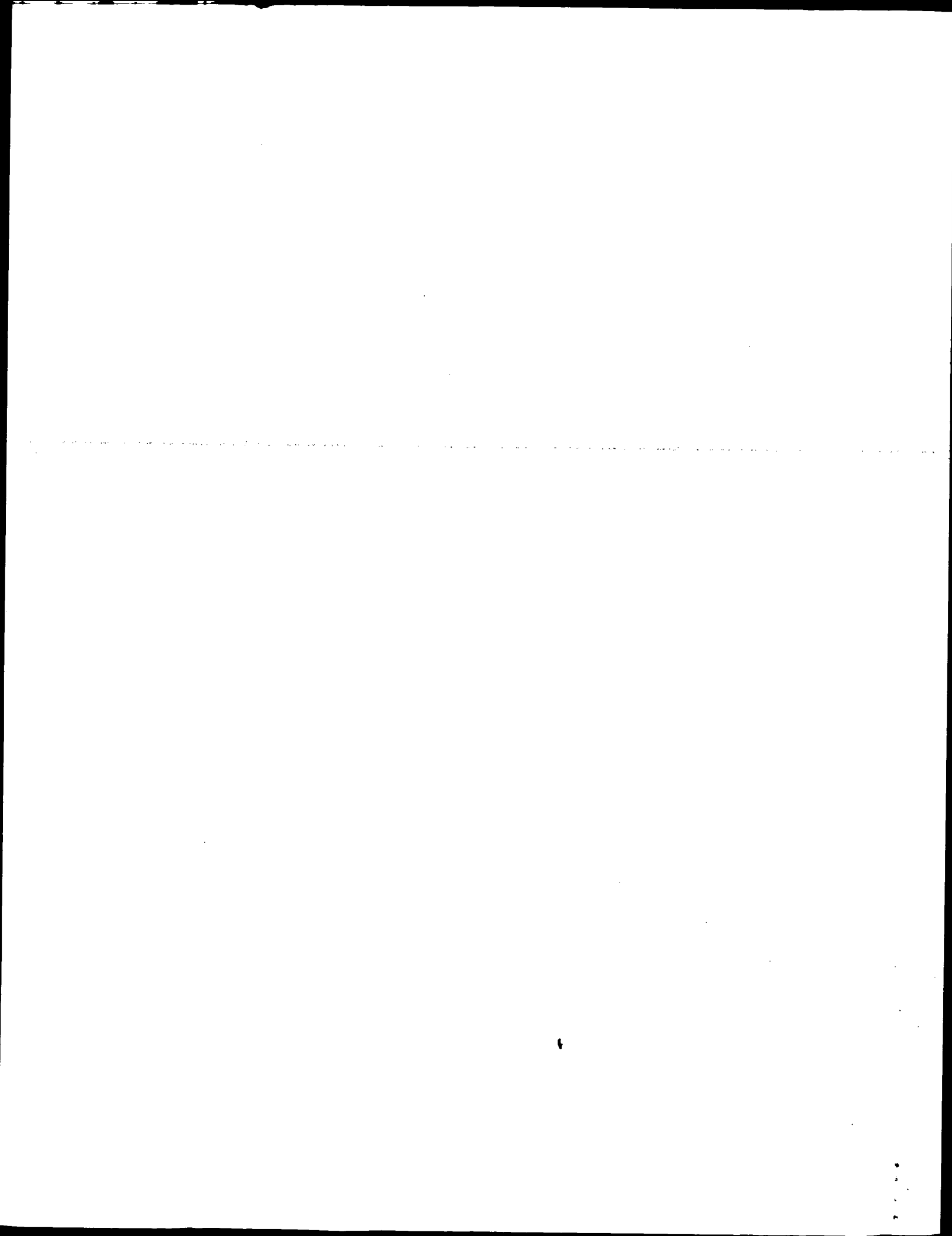
misc_feature /note="assembly_fragment"
16972..19133
misc_feature /note="assembly_fragment"
19234..22231
clone_end:SP6
vector_side:right
22332..26181
misc_feature /note="assembly_fragment"
26282..29449
misc_feature /note="assembly_fragment"
29550..33856
misc_feature /note="assembly_fragment"
33957..37947
misc_feature /note="assembly_fragment"
38048..41868
misc_feature /note="assembly_fragment"
41969..46321
misc_feature /note="assembly_fragment"
46422..50303
misc_feature /note="assembly_fragment"
50404..57011
misc_feature /note="assembly_fragment"
57112..62745
misc_feature /note="assembly_fragment"
62846..69054
misc_feature /note="assembly_fragment"
69155..75147
misc_feature /note="assembly_fragment"
75248..82842
misc_feature /note="assembly_fragment"
82943..90759
misc_feature /note="assembly_fragment"
90860..102877
misc_feature /note="assembly_fragment"
102978..115507
misc_feature /note="assembly_fragment"
115608..128821
misc_feature /note="assembly_fragment"
128922..138330
clone_end:T7
vector_side:right
138431..153959
misc_feature /note="assembly_fragment"

```

Query Match 2.7%; Score 26; DB 52; Length 177491;  
 Best Local Similarity 100.0%; Pred. No. 0.013;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 atttgatactgttatgaattgtaatt 427  
 |||||  
 Db 68504 ATTTGATACCTGTATGAATTGTAAT 68479

Search completed: November 4, 2000, 04:20:02  
 Job time: 22312 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:38:29 ; Search time 255.06 Seconds  
(without alignments)  
1413.927 Million cell updates/sec

Title: US-09-101-423A-5  
Perfect score: 960  
Sequence: 1 gaggggtgtggcacagt.....tcaaccaaccaaccacat 960

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 480022 seqs, 187831343 residues

Word size : 0

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	807	84.1	960	18	Metastasis inducin
2	26	2.7	1678	17	Probe for detectin
3	21	2.2	2830	18	Mouse receptor pro
4	21	2.2	2830	20	Mouse interleukin-
5	21	2.2	36901	20	Murine LOBO genomi
6	21	2.2	38886	20	Murine integrin sub
7	20	2.1	7958	19	Mouse Fas ligand-1
8	20	2.1	9058	19	Cellulose synthase
9	19	2.0	902	20	Nucleotide sequenc
10	19	2.0	6932	21	Target sequence #7
11	18	1.9	26	14	Antitumoural phosph
12	18	1.9	28	18	T93822

c 13	18	1.9	876	20	X57461	Rat U3 gene trap d
c 14	18	1.9	1002	20	X57454	Rat U3 gene trap d
c 15	18	1.9	1031	20	X57455	Rat U3 gene trap d
c 16	18	1.9	1595	21	297025	Human secreted pro
c 17	18	1.9	31096	18	V74370	Staphylococcus aur
c 18	18	1.9	37940	20	Z01026	Partial mouse PGI
c 19	18	1.9	235033	19	V57926	Hereditary haemoch
c 20	18	1.9	237326	19	V57903	Hereditary haemoch
c 21	17	1.8	136	19	X12018	Human biallelic po
c 22	17	1.8	233	20	V85141	EST clone H438. H
c 23	17	1.8	307	20	V90093	EST clone CY11. H
c 24	17	1.8	335	14	Q60081	Human brain expres
c 25	17	1.8	427	17	T34599	Probe for detectin
c 26	17	1.8	520	20	X07514	Human endogenous r
c 27	17	1.8	520	20	X07189	IDDM-associated hu
c 28	17	1.8	603	20	X07515	Human endogenous r
c 29	17	1.8	603	20	X07190	Human endogenous r
c 30	17	1.8	604	20	X07516	Human endogenous r
c 31	17	1.8	604	20	X07191	Human endogenous r
c 32	17	1.8	676	19	V40454	5' portion of H438
c 33	17	1.8	677	20	X04876	Human gamma synucl
c 34	17	1.8	703	21	Z80619	Human colon cancer
c 35	17	1.8	720	20	X23997	Human colon cancer
c 36	17	1.8	842	21	A02457	Human colon cancer
c 37	17	1.8	1048	7	M60313	Equine IFN-alpha-2
c 38	17	1.8	1079	21	A01687	Human colon cancer
c 39	17	1.8	1179	14	Q46868	Growth factor gene
c 40	17	1.8	1353	19	V65287	DNA encoding a S.
c 41	17	1.8	1532	20	X58810	Human leukaemia-as
c 42	17	1.8	1754	20	X07513	Human endogenous r
c 43	17	1.8	1754	20	X07188	IDDM-associated hu
c 44	17	1.8	1953	21	A26986	Streptococcus pneu
c 45	17	1.8	2148	13	Q27307	Rabbit endopeptida

ALIGNMENTS

RESULT 1	
T72789	
ID T72789 standard; DNA; 960 BP.	
XX	
AC T72789;	
XX	
DT 22-SEP-1997 (first entry)	
XX	
DE Metastasis inducing DNA C12.	
XX	
KW Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;	
KW osteopontin; ss:	
XX	
OS Homo sapiens.	
XX	
PN WO9725443-Al.	
XX	
PD 17-JUL-1997.	
XX	
PF 10-JAN-1997; 97WO-GB000074.	
XX	
PR 10-JAN-1996; 96GB-0000470.	
XX	
PA (UYLI-) UNIV LIVERPOOL.	
XX	
PI Barracough BR, Rudland PS;	
XX	
DR WPI; 1997-372878/34.	
XX	
PT New isolated metastasis-inducing DNA - used to develop products to	
PT identify and treat patients at risk from metastatic tumours	
XX	
PS Claim 12; Page 27; 38pp; English.	
XX	
CC Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12	

and C20 (772785-90) are entirely novel short stretches of human regulatory DNA capable of inducing metastasis. They were identified using a method for detecting Met-DNA that involves transferring human DNA from a malignant, metastatic cancer cell (in this case, breast cancer) into a cell line (pref. rat Rama 37) that produces only benign, non-metastasing tumours when injected into a syngeneic animal, injecting those transformed cells into a syngeneic animal, selecting those animals having metastasing tumours, and recovering the Met-DNA from them. The isolated Met-rDNAs can be used to develop products to identify and treat patients at risk from metastatic tumours.

Sequence 960 BP; 256 A; 193 C; 233 G; 278 T; 0 other:

Db	781	gcaaaagggggggtgacacatagtgtcattatgacacacatctcgtgtgattaaagggggagtcag	840
Qy	841	ttccttggggggaagtttgattcttcgtgaacgatatctaaattctctccctgttgctt	900
Db	841	ttccttgggggaagtttgattcttcgtgaacgatatctaaattctctccctgttgctt	900
Qy	901	cgtcttgtgacaaagcagactgtataccccaatggacatcaacccaaccaaacat	960
Db	901	cgtcttgtgacaaagcagactgtataccccaatggacatcaacccaaccaaacat	960

## RESULT 2

RESCUE 2  
T34611  
ID T34611 standard; cDNA; 1678 BP.

XX	
AC	T34611;
XX	
XX	02-JAN-1997 (first entry)
DT	
XX	
DE	Probe for detecting pathology in peripheral nervous system.

XX Nucleic acids associated with damaged or regenerating nerve cells  
PT for diagnostic or therapeutic use.  
XX  
XX  
PS Claim 1: Page 35-37: 54pp: German.

Nucleic acids described in T34357-T34616 are useful as probes for detecting traumatic and pathological changes in the peripheral nervous system. They may also be used to identify hybridizing sequences from an ischial nerve cDNA library, these sequences encode proteins expressed in damaged and/or regenerating nerve cells. They can be used to transfect cells which then express the protein which can then be harvested and studied. The nucleic acids are preferably selected from NtII-11 which corresponds to plasminogen cDNA with a 234 bp N-terminal extension and is expressed in glial cells of the central nervous system and Schwann cells of the peripheral nervous system; NtII-1, which codes for a protein having 79% homology with human CDC4L in three domains and may be involved in regulating Schwann cell growth as well as being therapeutically useful for regenerating nerve cells; and CRII-7 which codes for a protein which is similar to the macrophage-specific cysteine-rich mouse protein MS2 and a metalloprotease and which inhibits interaction between Schwann cells and the extracellular matrix and is useful for the therapy of nerve damage by facilitating regeneration of damaged axons. This sequence is designated NtI-2a.

Sequence 1678 B: 427 A: 380 C: 443 G: 427 T: 1 other: 1

QY 839 agtctctggggcaagttgtgatttt 864  
 |||||  
 Db 1229 agtctctggggcaagttgtgatttt 1254

RESULT 3  
 T88775/c  
 ID T88775 standard; cDNA; 2830 BP.  
 XX  
 AC T88775;  
 XX  
 DT 22-FEB-1998 (first entry)  
 XX  
 DE Mouse receptor protein 2f1 cDNA clone mu2f1.

XX 2f1; receptor; prostaglandin; inflammation; arthritis; allergy;  
 KW asthma; inflammatory bowel disease; fever; diabetes;  
 KW dysmenorrhea; therapy; mouse; ss.  
 XX  
 OS Mus musculus.

XX Key Location/Qualifiers  
 FH CDS 381..1994  
 FT /\*tag= a  
 FT sig\_peptide 381..434  
 FT /\*tag= b  
 FT mat\_peptide 435..1991  
 FT /\*tag= c

XX W09731010-A1.  
 XX 28-AUG-1997.  
 XX 06-FEB-1997; 97WO-US01697.  
 XX 21-FEB-1996; 96US-0604333.  
 XX (IMMV ) IMMUNEX CORP.  
 XX  
 XX Parnet P, Sims JE;  
 XX WPI; 1997-435081/40.  
 XX P-PSDB; W31908.

XX DNA encoding 2f1 receptors that inhibit prostaglandin synthesis -  
 PT for treatment of inflammatory diseases, e.g. arthritis, allergy,  
 PT asthma, etc.  
 XX  
 XX Example 2; Page 36-38; 49pp; English.

XX cDNA clone mu2f1 encodes a novel mouse receptor protein designated  
 CC 2f1 (see W31908) that is a member of the interleukin-1 receptor  
 CC (IL-1R) family. It was isolated from a murine EL4 6.1 cDNA library  
 CC using a probe based on human 2f1 cDNA (see T88774). The isolated  
 CC cDNA can be used for production of recombinant 2f1, especially  
 CC soluble 2f1, in transfected host cells for use in claimed methods  
 CC of inhibiting prostaglandin synthesis in a mammal, and of reducing  
 CC inflammation in a mammal (e.g. for treating arthritis, allergy,  
 CC asthma, inflammatory bowel disease, fever, diabetes, dysmenorrhea,  
 CC etc.). Since 2f1 can inhibit NF-kappaB activity, it may also be  
 CC used to reduce expression of human immunodeficiency virus in  
 CC infected cells (not claimed). Fragments of DNA are useful as probes  
 CC to isolate related sequences, in hybridisation assays and as  
 CC sense or antisense molecules to block expression of 2f1 protein.  
 XX  
 XX Sequence 2830 BP; 830 A; 594 C; 669 G; 737 T; 0 other;

Query Match 2.2%; Score 21; DB 18; Length 2830;  
 Best Local Similarity 100.0%; Pred. No. 0.73;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 ccagcgggtctcaaccttct 316  
 |||||  
 Db 2433 CCAGCGGTCTCAACCTTCT 2413

RESULT 4  
 X87653/c  
 ID X87653 standard; DNA; 2830 BP.  
 XX  
 AC X87653;  
 XX  
 DT 26-OCT-1999 (first entry)  
 XX  
 DE Mouse interleukin-1 receptor related protein 1 DNA.

XX Interleukin-1 receptor related protein 1; IL-1Rpl; mouse;  
 KW interleukin-18 receptor; IL-18; signal transduction; inflammation;  
 KW antiinflammatory; autoimmune disease; therapy; ss.  
 XX  
 OS Mus musculus.

XX Key Location/Qualifiers  
 FH sig\_peptide 1..54  
 FT /\*tag= a  
 FT mat\_peptide 55..1991  
 FT /\*tag= b

XX W09937772-A1.  
 XX 29-JUL-1999.  
 XX 22-JAN-1999; 99WO-US01419.  
 XX 28-JUL-1998; 98US-0094469.  
 XX 23-JAN-1998; 98US-0072301.  
 XX 20-MAR-1998; 98US-0078835.  
 XX (IMMV ) IMMUNEX CORP.

XX Born TL, Sims JE;  
 XX WPI; 1999-493952/41.  
 XX P-PSDB; Y06615.

XX Interleukin-18 receptor complex useful for treating inflammatory  
 PT and/or autoimmune diseases attributable to IL-18 signaling  
 XX  
 XX Claim 1; Page 48-51; 56pp; English.

XX This DNA sequence codes for mouse interleukin-1 receptor related  
 CC protein 1 (IL-1Rpl, see Y06615), a subunit of the interleukin-18  
 CC (IL-18) receptor complex. Polypeptides that function as IL-18  
 CC receptors are disclosed. Such receptors are multimeric and include  
 CC at least one AcPL polypeptide (see Y06614) and at least one  
 CC IL-1Rpl polypeptide, or soluble fragments thereof. The dimeric  
 CC IL-18 receptor complexes can include IL-1Rpl and AcPL as proteins  
 CC coexpressed in the same cell, or as IL-1Rpl linked to AcPL as  
 CC receptor subunits, preferably linked via covalent linkages.  
 CC Coexpression of AcPL and IL-1Rpl results in a dramatic enhancement  
 CC of NFkB activity in cells stimulated with IL-18. The receptors  
 CC mediate signal transduction and can be used to inhibit IL-18 and  
 CC ameliorate any inflammatory and/or autoimmune diseases attributable  
 CC to IL-18 signal transduction.

XX Sequence 2830 BP; 830 A; 594 C; 669 G; 737 T; 0 other;

Query Match 2.2%; Score 21; DB 20; Length 2830;  
 Best Local Similarity 100.0%; Pred. No. 0.73;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 ccagcgggtctcaaccttct 316  
 |||||

Db 2433 CCAGCGTTCTCAAGCTTCT 2413

RESULT 5

223892 ID 223892 standard; DNA; 36901 BP.

XX AC 223892;

XX DT 25-JAN-2000 (first entry)

XX DE Murine LOBO genomic DNA fragment 2.

XX KW LOBO; long bones; bone development; bone extension; skull; osteopathic;  
diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
XX KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.

XX OS Mus musculus.

XX PN WO9950284-A2.

XX PD 07-OCT-1999.

XX PF 26-MAR-1999; 99WO-EP02055.

XX PR 27-MAR-1998; 98DE-1013799.

XX PA (ROSE/) ROSENTHAL A.

XX PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;

XX DR WPI; 1999-601320/51.

XX CC Nucleic acids encoding proteins which influence bone development,  
useful for treating and studying bone disorders -

XX PS Example 3; Page 97-117; 391pp; German.

XX CC This invention describes novel nucleic acids (I; designated LOBO (long  
bones)) encoding proteins influencing bone development in mammals. The  
proteins of the invention reduce and/or inactivate bone extension (i.e.  
development), with exception of the skull and have osteopathic activity.  
XX CC The nucleic acid molecules, proteins and antibodies can be used in  
diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
and nucleic acid molecules, etc. are useful for production of transgenic  
XX CC animals, especially a transgenic mouse for the study of diseases  
associated with bone development, e.g. spondyloepiphyseal dysplasia and  
XX CC achondroplasia. This sequence encodes the murine LOBO protein described  
in the method of the invention.

XX SQ Sequence 36901 BP; 9597 A; 9409 C; 8973 G; 8922 T; 0 other;

Query Match

Best Local Similarity 2.28; Score 21; DB 20; Length 36901;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 301 gggtctcaaccttctctgac 321

Db 17559 gggtctcaaccttctctgac 17579

RESULT 6

223897

XX ID 223897 standard; DNA; 38886 BP.

XX AC 223897;

XX DT 25-JAN-2000 (first entry)

XX DE Murine LOBO homologue genomic DNA fragment 3.

XX KW LOBO; long bones; bone development; bone extension; skull; osteopathic;

KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.

XX OS Mus musculus.

XX PN WO9950284-A2.

XX PD 07-OCT-1999.

XX PF 26-MAR-1999; 99WO-EP02055.

XX PR 27-MAR-1998; 98DE-1013799.

XX PA (ROSE/) ROSENTHAL A.

XX PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;

XX DR WPI; 1999-601320/51.

XX CC Nucleic acids encoding proteins which influence bone development,  
useful for treating and studying bone disorders -

XX PS Example 3; Page 189-211; 391pp; German.

XX CC This invention describes novel nucleic acids (I; designated LOBO (long  
bones)) encoding proteins influencing bone development in mammals. The  
proteins of the invention reduce and/or inactivate bone extension (i.e.  
development), with exception of the skull and have osteopathic activity.  
XX CC The nucleic acid molecules, proteins and antibodies can be used in  
diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods  
and nucleic acid molecules, etc. are useful for production of transgenic  
XX CC animals, especially a transgenic mouse for the study of diseases  
associated with bone development, e.g. spondyloepiphyseal dysplasia and  
XX CC achondroplasia. This sequence encodes the murine LOBO protein described  
in the method of the invention.

XX SQ Sequence 38886 BP; 10087 A; 9920 C; 9356 G; 9523 T; 0 other;

Query Match

Best Local Similarity 2.28; Score 21; DB 20; Length 38886;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 301 gggtctcaaccttctctgac 321

Db 19543 gggtctcaaccttctctgac 19563

RESULT 7

234835

XX ID 234835 standard; DNA; 7958 BP.

XX AC 234835;

XX DT 15-FEB-2000 (first entry)

XX DE Mouse integrin subunit alpha-10 gene.

XX KW Integrin alpha-10; ISa10; mouse; trauma; rheumatoid arthritis;  
osteoarthritis; osteoarthritis; cancer; atherosclerosis;  
XX KW inflammation; therapy; cartilage; chondrocyte; osteoblast;  
fibroblast; vaccine; marker; ss.

XX OS Mus musculus.

XX FH Key Location/Qualifiers

XX FT CDS 2672..7580

XX FT /tag= a

XX FT /note= "contains introns"

XX FT exon 2672..2723

XX FT /tag= b

XX FT /number= 1

XX FT intron 2724..4112







```
FT      /*tag= f
FT      /transl_except= (pos: 5634..5636, aa: Xaa)
FT      /transl_except= (pos: 5637..5638, aa: Xaa)
FT      /note= "CpsIIJ; Xaa is an unspecified amino acid;
FT      encodes Y68975; nucleotides 5328-5330
FT      encode an amino acid that is not included
FT      in the protein sequence"
FT      CDS
FT      6156..6990
FT      /*tag= g
FT      /note= "CpsIIK; encodes Y68976; no termination codon
FT      given"
FT      PN
FT      WO200005378-A2.
FT      PD
FT      03-FEB-2000.
FT      XX
FT      19-JUL-1999; 99WO-NL00460.
FT      XX
FT      22-JUL-1998; 98EP-0202465.
FT      PR
FT      22-JUL-1998; 98EP-0202467.
FT      XX
FT      (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
FT      PA
FT      Smith HE;
FT      XX
FT      WPI; 2000-195104/17.
FT      DR
FT      P-PSDB; Y68970, Y68971, Y68972, Y68974, Y68975, Y68976.
FT      XX
FT      New nucleic acid containing the capsular gene cluster of Streptococcus
FT      suis, used for serotype-specific detection and to generate antigens or
FT      mutants for vaccination
FT      PT
FT      Claim 5; Fig 4; 144pp; English.
FT      XX
FT      The present sequence represents the capsular gene cluster of
FT      Streptococcus suis serotype 1. The genes in this cluster are involved
FT      in polysaccharide biosynthesis of capsular components and antigens. The
FT      proteins have glycosyltransferase activities (CpsIE, CpsIG, CpsII,
FT      CpsIIJ, CpsIIK) and CP polymerase activities (CpsIIH). The capsule confers
FT      bacterium resistance to complement-mediated opsonophagocytosis. The
FT      gene cluster is used as a source of probes and primers for
FT      serotype-specific detection of S. suis and is also useful for
FT      recombinant production of the proteins. The proteins are then useful
FT      for producing antigens that can be used in vaccines, for controlling
FT      or eradicating a Streptococcal disease, in humans or animals,
FT      e.g. against S. suis in pigs.
FT      CC
FT      Sequence 6992 BP; 2513 A; 788 C; 1214 G; 2477 T; 0 other;
FT      XX
FT      Query Match 2.0%; Score 19; DB 21; Length 6992;
FT      Best Local Similarity 100.0%; Pred. No. 6.9;
FT      Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT      QY 425 aatataaataatttgaag 443
FT      Db 5541 aatataaataatttgaag 5559
FT      |||||||
FT      RESULT 11
FT      Q44014/c
FT      ID Q44014 standard; DNA; 26 BP.
FT      XX
FT      Q44014;
FT      XX
FT      28-OCT-1993 (first entry)
FT      DT
FT      Target sequence #7.
FT      DE
FT      Purine; pyrimidine; tracts; therapeutic; diagnostic;
FT      XX
FT      control; gene expression; mRNA synthesis suppression; ds.
FT      KW
FT      Synthetic.
FT      XX
FT      OS
```

```
XX      WO9312230-A.
PN
XX      24-JUN-1993.
PD
XX      11-DEC-1992; 92WO-US10792.
PF
XX      13-DEC-1991; 91US-0808452.
PR
XX      21-JAN-1992; 92US-0826934.
PT
XX      (STRI ) SRI INT.
PS
XX      Jayasena SD, Johnston BH;
XX      WPI; 1993-214172/26.
XX      New oligo:nucleotide(s) forming triple helix with target nucleic
XX      acid - contain purine and pyrimidine tracts in specific
XX      orientations, useful therapeutically or diagnostically e.g. for
XX      inactivating HIV RNA, etc.
XX      Example; Fig 14a; 101pp; English.
XX      The sequence is that of the target sequence #7 which was used in an
XX      experiment to determine the in vitro cleavage of target duplexes to
XX      evaluate the lengths of purine and pyrimidine tracts which are
XX      useful in obtaining oligonucleotides capable of triple helix
XX      formation with target nucleic acids. The complementary strand
XX      overhangs the 3' end by the sequence CRAG and the sense strand
XX      overhangs the complementary strand by the sequence AATT.
XX      CC
XX      Sequence 26 BP; 2 A; 1 C; 11 G; 12 T; 0 other;
XX      SQ
XX      Query Match 1.9%; Score 18; DB 14; Length 26;
XX      Best Local Similarity 100.0%; Pred. No. 23;
XX      Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX      QY 942 caaccacaccacacacac 959
XX      Db 26 CAACCAACCAACCAACCA 9
XX      |||||||
XX      RESULT 12
XX      T93822/c
XX      ID T93822 standard; DNA; 28 BP.
XX      AC
XX      T93822;
XX      XX
XX      24-FEB-1998 (first entry)
XX      DT
XX      Antitumoral phosphodiester oligonucleotide 12 with cytotoxic activity.
XX      DE
XX      Phosphodiester; selective binding; cell viability; growth;
XX      tumoural cell line; cytotoxic activity; tumour cell; lymphoma;
XX      lymphoblastic tumour; ss.
XX      KW
XX      Synthetic.
XX      OS
XX      Key Location/Qualifiers
XX      modified_base 1..28
XX      /tag= a
XX      /note= "phosphodiester oligonucleotide"
XX      FT
XX      WO9720924-A1.
XX      PN
XX      12-JUN-1997.
XX      PD
XX      04-DEC-1996; 96WO-EP05388.
XX      PF
XX      04-DEC-1995; 95IT-OMI2539.
XX      PR
XX      (SAIC-) SAICOM SRL.
XX      PA
```

XX Quadrifoglio F, Scaggiante B;  
 XX WPI; 1997-319771/29.  
 XX  
 PT New phosphodiesteric oligonucleotide(s) - which exert a specific  
 PT and selective cytotoxic effect on tumour cells, for treating both  
 PT solid and liquid tumours  
 XX  
 PS Claim 10; Page 6; 38pp; English.  
 XX  
 CC Novel phosphodiesteric oligonucleotides T93811-27 are based on the  
 CC generic formula, in the 3'-5' or 5'-3' direction:  
 CC (Gara'a'-(Gbrb')b'-(Gctc')c'-(Gdtd')d'-(GeTe')e'-(Gfif')f'-  
 CC (G-gtg')g'-N', where:  
 CC N and N' = T or G, equal or different from each other;  
 CC x = 0-8, equal or different from each other;  
 CC a, b, c, d, e, f, and g = 0-10, equal or different from each other;  
 CC a', b', c', d', e', f', and g' = 0-30, equal or different from each  
 CC other;  
 CC a'', b'', c'', d'', e'', f'', and g'' = 1-16, equal or different from  
 CC each other;  
 CC The oligonucleotides are believed to selectively bind and sequester  
 CC some proteins which are essential to the viability and growth of  
 CC tumour cell line. They have specific and selective cytotoxic activity  
 CC against tumour cells, and can be used for treating tumours of the liquid  
 CC type, in particular of lymphoblastic origin, and of solid type, in  
 CC particular lymphomas. The present phosphodiester oligonucleotide, at a  
 CC concentration of 15 micromolar, reduced growth of CCRF-CEM tumoural cells  
 CC by 70%, which is detectable 48 hours after administration.  
 XX  
 SQ Sequence 28 BP; 0 A; 0 C; 14 G; 14 T; 0 other;

Query Match 1.9%; Score 18; DB 18; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 942 caaccacccaaccaacca 959  
 Db 26 CAACCAACCAACCAACCA 9

RESULT 13  
 X57461/c  
 ID X57461 standard; DNA; 876 BP.  
 XX  
 AC X57461;  
 XX  
 DT 24-JUL-1999 (first entry)  
 XX  
 DE Rat U3 gene trap derived nucleic acid L24 26-2A/re.

Gene trap; rat; cellular gene; viral infection; cell survival; cancer;  
 tumour progression; suppression; identification; viral growth;  
 tumour suppressor; prevention; screening; therapeutic agent; ss.

OS Rattus norvegicus.

XX WO9919481-A2.

XX 22-APR-1999.

XX 08-OCT-1998; 98WO-US21276.

XX 10-OCT-1997; 97US-0062021.

XX (UYVA-) UNIV VANDERBILT.

XX Dubois RN, Organ EL, Rubin DH;

XX WPI; 1999-326546/27.

XX

PT Nucleic acid encoding tumor suppressors and products required for  
 PT viral infection.  
 XX  
 XX Claim 1; Page 73; 94pp; English.  
 XX  
 CC This invention describes novel rat-derived nucleic acid fragments from  
 CC cellular genes that are necessary for viral infection but not for cell  
 CC survival, or that suppress tumour progression. The products of the  
 CC invention (X57371-X57497) can be used in methods of identifying cellular  
 CC genes necessary for viral growth and cellular genes that function as  
 CC tumour suppressors and for reducing or preventing such infections or  
 CC cancer. They may also be used in screening for potential therapeutic  
 CC agents. These sequences can be targeted without significant side effects  
 CC (contrast targeting genes essential for viral growth).  
 XX  
 SQ Sequence 876 BP; 221 A; 166 C; 172 G; 306 T; 11 other;

Query Match 1.9%; Score 18; DB 20; Length 876;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 942 caaccacccaaccaacca 959  
 Db 602 CAACCAACCAACCAACCA 585

RESULT 14  
 X57454/c  
 ID X57454 standard; DNA; 1002 BP.

XX AC X57454;

XX DT 24-JUL-1999 (first entry)

DE Rat U3 gene trap derived nucleic acid L22-5A1/re.

Gene trap; rat; cellular gene; viral infection; cell survival; cancer;  
 tumour progression; suppression; identification; viral growth;  
 tumour suppressor; prevention; screening; therapeutic agent; ss.

OS Rattus norvegicus.

XX WO9919481-A2.

XX 22-APR-1999.

XX 08-OCT-1998; 98WO-US21276.

XX 10-OCT-1997; 97US-0062021.

XX (UYVA-) UNIV VANDERBILT.

XX Dubois RN, Organ EL, Rubin DH;

XX WPI; 1999-326545/27.

PT Nucleic acid encoding tumor suppressors and products required for  
 PT viral infection.

XX Claim 1; Page 75; 94pp; English.

CC This invention describes novel rat-derived nucleic acid fragments from  
 CC cellular genes that are necessary for viral infection but not for cell  
 CC survival, or that suppress tumour progression. The products of the  
 CC invention (X57371-X57497) can be used in methods of identifying cellular  
 CC genes necessary for viral growth and cellular genes that function as  
 CC tumour suppressors and for reducing or preventing such infections or  
 CC cancer. They may also be used in screening for potential therapeutic  
 CC agents. These sequences can be targeted without significant side effects  
 CC (contrast targeting genes essential for viral growth).

XX Sequence 1002 BP; 259 A; 173 C; 207 G; 344 T; 19 other;

Query Match 1.9%; Score 18; DB 20; Length 1002;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 942 caaccacacacacacacac 959  
| | | | | | | | | | | | | | | |  
DB 735 CAACCAACCAACCAACCA 718

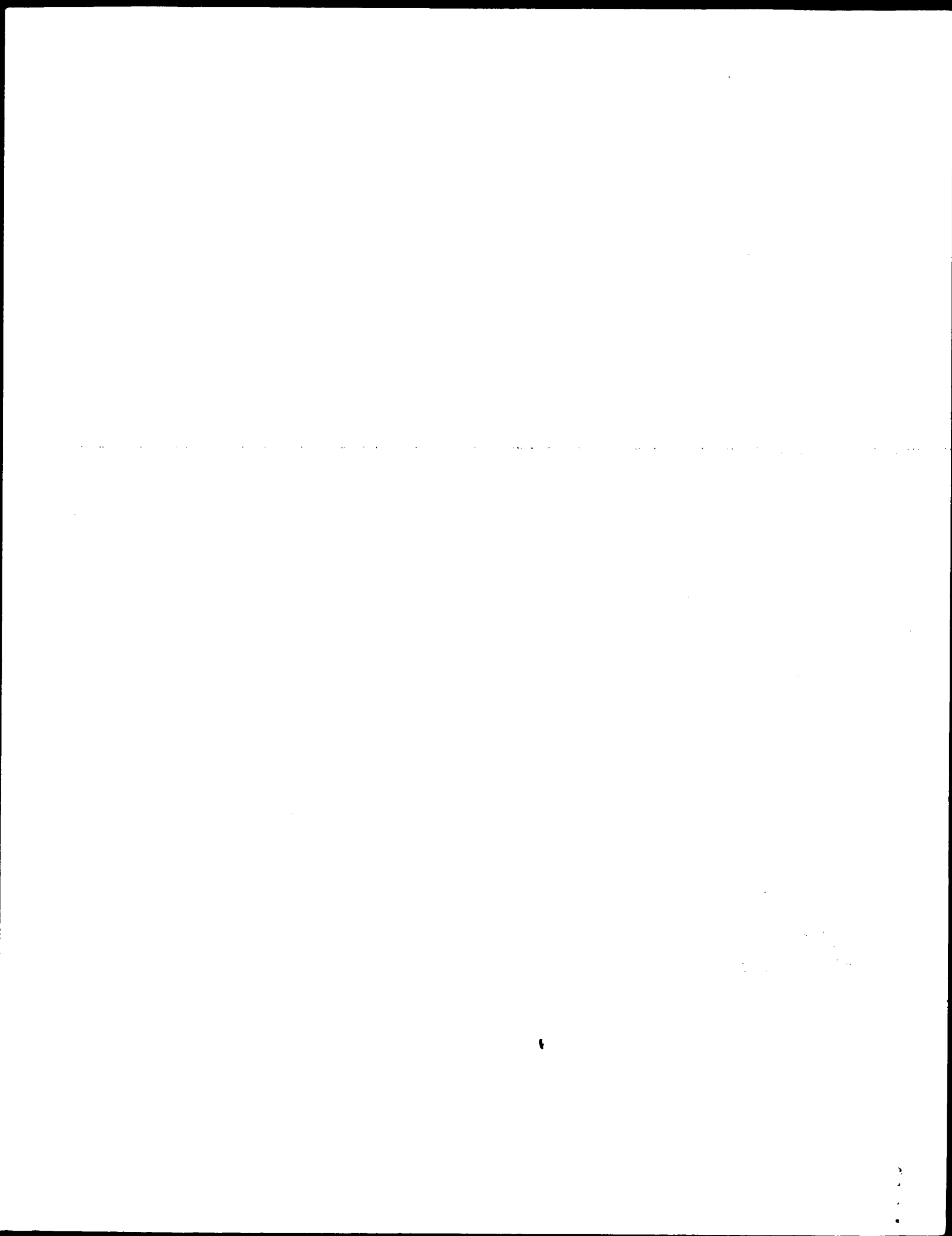
RESULT 15  
X57455/C  
ID X57455 standard; DNA; 1031 BP.  
XX AC X57455;  
XX 24-JUL-1999 (first entry)  
XX DE Rat U3 gene trap derived nucleic acid L24\_3\_2B/RE.  
XX KW Gene trap; rat; cellular gene; viral infection; cell survival; cancer;  
KW tumour progression; suppression; identification; viral growth;  
KW tumour suppressor; prevention; screening; therapeutic agent; ss.  
OS Rattus norvegicus.  
XX PN WO9919481-A2.  
XX PD 22-APR-1999.  
XX PF 08-OCT-1998; 98WO-US21276.  
XX PR 10-OCT-1997; 97US-0062021.  
XX PA (UYVA-) UNIV VANDERBILT.  
XX PI DuBois RN, Organ<sup>EL</sup>, Rubin DH;  
XX DR WPI; 1999-326546/27.  
XX PT Nucleic acid encoding tumor suppressors and products required for  
XX viral infection

Claim 1; Page 76; 94pp; English.  
XX CC This invention describes novel rat-derived nucleic acid fragments from  
XX cellular genes that are necessary for viral infection but not for cell  
XX survival, or that suppress tumour progression. The products of the  
XX invention (X57371-X57497) can be used in methods of identifying cellular  
XX genes necessary for viral growth and cellular genes that function as  
XX tumour suppressors and for reducing or preventing such infections or  
XX cancer. They may also be used in screening for potential therapeutic  
XX agents. These sequences can be targeted without significant side effects  
XX (contrast targeting genes essential for viral growth).  
XX SQ Sequence 1031 BP; 273 A; 180 C; 204 G; 352 T; 22 other;

Query Match 1.9%; Score 18; DB 20; Length 1031;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 942 caaccacacacacacacac 959  
| | | | | | | | | | | | | | | |  
DB 757 CAACCAACCAACCAACCA 740

Search completed: November 4, 2000, 03:39:12  
Job time: 11216 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:33:06 ; Search time 140.17 Seconds  
(without alignments)  
1035.826 Million cell updates/sec

Title: US-09-101-423A-5  
Perfect score: 960  
Sequence: 1 gaggaggtgtgtgcacagt.....tcaaccaaccaaccaaccat 960

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 262060 seqs, 75620727 residues

Word size : 0

total number of hits satisfying	chosen parameters:	524120
---------------------------------	--------------------	--------

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

```
Database :
Issued_Patents_NA:
1: /c3n2_6/prodata/2/ina/5A_COMB.seq:
2: /c3n2_6/prodata/2/ina/5B_COMB.seq:
3: /c3n2_6/prodata/2/ina/5C_COMB.seq:
4: /c3n2_6/prodata/2/ina/5D_COMB.seq:
5: /c3n2_6/prodata/2/ina/6_COMB.seq:
6: /c3n2_6/prodata/2/ina/PCUTS_COMB.seq:
7: /c3n2_6/prodata/2/ina/PCUTS_BACKFILE1.seq:
```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	21	2.2	2830	2	US-08-604-333-3	Sequence 3, Appl1
C 2	21	2.2	2830	5	US-09-110-618-3	Sequence 3, Appl1
C 3	18	1.9	26	6	PCT-US92-10792-42	Sequence 42, Appl
C 4	18	1.9	26	6	PCT-US92-10792-43	Sequence 43, Appl
C 5	18	1.9	3252	6	PCT-US96-05320A-714	Sequence 714, App
C 6	18	1.9	246240	3	US-08-724-394A-20	Sequence 20, Appl
C 7	18	1.9	246240	3	US-08-724-394A-21	Sequence 21, Appl
C 8	18	1.9	246240	3	US-08-724-394A-22	Sequence 22, Appl
C 9	17	1.8	1179	3	US-08-465-794-4	Sequence 4, Appl1
C 10	17	1.8	1179	5	US-09-049-813-4	Sequence 4, Appl1
C 11	17	1.8	2121	4	US-08-633-879C-17	Sequence 17, Appl
C 12	17	1.8	2148	6	PCT-US92-00731-12	Sequence 12, Appl
C 13	17	1.8	5115	5	US-08-348-518C-3	Sequence 3, Appl1
C 14	17	1.8	5115	5	US-08-476-509B-3	Sequence 3, Appl1
C 15	17	1.8	8252	1	US-08-046-585-15	Sequence 15, Appl
C 16	17	1.8	8252	1	US-08-393-703-15	Sequence 15, Appl
C 17	17	1.8	8252	6	PCT-US93-11721-15	Sequence 15, Appl
C 18	17	1.8	16442	5	US-08-781-891-208	Sequence 208, App
C 19	17	1.8	51259	5	US-08-781-891-209	Sequence 209, App
C 20	16	1.7	40	1	US-08-199-507B-43	Sequence 43, Appl
C 21	16	1.7	40	2	US-08-441-828-43	Sequence 43, Appl
C 22	16	1.7	152	1	US-08-112-816-3	Sequence 3, Appl1
C 23	16	1.7	239	1	US-08-248-474-20	Sequence 20, Appl
C 24	16	1.7	239	5	US-08-566-849-20	Sequence 20, Appl
C 25	16	1.7	471	1	US-08-112-816-1	Sequence 1, Appl1
C 26	16	1.7	775	5	US-09-124-476-1	Sequence 1, Appl1

27	16	1.7	786	6	PCT-US95-08295-1	Sequence 1, Appl
28	16	1.7	926	1	US-08-106-507-1	Sequence 1, Appl
29	16	1.7	957	4	US-08-975-316-81	Sequence 81, Appl
30	16	1.7	959	3	US-08-713-000-5	Sequence 5, Appl
31	16	1.7	959	4	US-08-975-316-5	Sequence 5, Appl
32	16	1.7	1016	1	US-08-109-391A-3	Sequence 3, Appl
33	16	1.7	1016	1	US-08-459-039A-3	Sequence 3, Appl
34	16	1.7	1016	4	US-08-460-428A-3	Sequence 3, Appl
35	16	1.7	1016	5	US-08-458-860A-3	Sequence 5, Appl
36	16	1.7	1047	2	US-08-843-521-5	Sequence 2, Appl
37	16	1.7	1047	3	US-08-953-041-1	Sequence 1, Appl
38	16	1.7	1468	5	US-09-215-131-2	Sequence 2, Appl
39	16	1.7	1468	5	US-09-079-981-2	Sequence 2, Appl
40	16	1.7	1468	5	US-09-222-734-2	Sequence 2, Appl
41	16	1.7	1500	1	US-07-914-281-3	Sequence 3, Appl
42	16	1.7	1500	1	US-08-393-246-3	Sequence 3, Appl
43	16	1.7	1500	2	US-08-525-058A-3	Sequence 3, Appl
44	16	1.7	1500	4	US-08-696-731-3	Sequence 2, Appl
45	16	1.7	1500	6	PCT-US91-00399-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-604-333-3/c  
; Sequence 3, Application US/08604333  
; Patent No. 5776731  
; GENERAL INFORMATION:  
; APPLICANT: Parinet, Patricia et al.  
; TITLE OF INVENTION: Receptor Designated 2f1  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kathryn A. Anderson, Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple 7.1  
; SOFTWARE: Microsoft Word, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/604,333  
; FILING DATE: 21-FEB-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Anderson, Kathryn A.  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2619  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; TELEX: 756822  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2830 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NC  
; IMMEDIATE SOURCE:  
; CLONE: mu2f1  
; FEATURE:  
; NAME/KEY: C1S  
; LOCATION: 381..1994  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 435..1991

FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 381..434  
US-08-604-333-3

Query Match 2.28; Score 21; DB 2; Length 2830;  
Best Local Similarity 100.0%; Pred. No. 0.26;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 ccagcgggttcacaccttct 316  
|||||  
Db 2433 CCAGCGGTTCACACCTTCT 2413

## RESULT 2

US-09-110-618-3/c  
; Sequence 3, Application US/09110618  
; Patent No. 6090918  
; GENERAL INFORMATION:  
; APPLICANT: Parnet, Patricia et al.  
; TITLE OF INVENTION: Receptor Designated 2F1  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kathryn A. Anderson, Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple 7.1  
; SOFTWARE: Microsoft Word, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/110.618  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/604.333  
; FILING DATE: 21-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Anderson, Kathryn A.  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2619  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; TELEX: 756822  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2830 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: mu2F1  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 381..1994  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 435..1991  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 381..434  
; US-09-110-618-3

Query Match 2.28; Score 21; DB 5; Length 2830;  
Best Local Similarity 100.0%; Pred. No. 0.26;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 296 ccagcgggttcacaccttct 316  
|||||  
Db 2433 CCAGCGGTTCACACCTTCT 2413

## RESULT 3

PCT-US92-10792-42/c  
; Sequence 42, Application PC/TUS9210792  
; GENERAL INFORMATION:  
; APPLICANT: Jayasena, Sumedha D.  
; APPLICANT: Johnston, Brian H.  
; TITLE OF INVENTION: Triple Helix Formation at  
; TITLE OF INVENTION: (PuNpYn)-(PuNpYn) Tracts  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: S3I International  
; STREET: 333 Ravenswood Avenue  
; CITY: Menlo Park  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/10792  
; FILING DATE: 19921211  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/826,934  
; FILING DATE: 21-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/808,452  
; FILING DATE: 13-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: P-3141  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 859-4550  
; TELEFAX: (415) 859-3880  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: FIGURE 14A, TARGET SEQUENCE #7  
; PCT-US92-10792-42

Query Match 1.9%; Score 18; DB 6; Length 26;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 942 caaccaaccaaccaacca 959  
|||||  
Db 26 CAACCAACCAACCAACCA 9

## RESULT 4

PCT-US92-10792-43  
; Sequence 43, Application PC/TUS9210792  
; GENERAL INFORMATION:

```

APPLICANT: Jayasena, Sumedha D.
APPLICANT: Johnston, Brian H.
TITLE OF INVENTION: Triple Helix Formation at
TITLE OF INVENTION: (PUNPYN)-(PUNPYN) Tracts
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: SRI International
STREET: 333 Ravenswood Avenue
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10792
FILING DATE: 19921211
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/826,934
FILING DATE: 21-JAN-1992
PRIOR APPLICATION NUMBER: US 07/808,452
FILING DATE: 13-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: P-3141
TELEPHONE INFORMATION:
TELEPHONE: (415) 859-4550
TELEFAX: (415) 859-3880
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: FIGURE 14A, TARGET SEQUENC
PCT-US92-10792-43

Query Match
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 18; Conservative 0; Mismatches 0;

QY 942 caacacacacacacacacac 959
Db 3 CAACCAACCAACCAACCA 20
|||||
|||||

RESULT 5
PCT-US96-05320A-714
Sequence 714, Application PC/TUS9605320A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences
APPLICANT: 9410 Key West Avenue
APPLICANT: Rockville, MD 20850
APPLICANT: United States of America
APPLICANT: Johns Hopkins University
APPLICANT: 720 Rutland Avenue
APPLICANT: Baltimore, MD 21205
APPLICANT: United States of America
APPLICANT: Mark D. Adams
APPLICANT: Owen White
APPLICANT: Hamilton O. Smith
APPLICANT: J. Craig Venter

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONTIG"  
US-08-724-394A-20

Query Match 1.9%; Score 18; DB 3; Length 246240;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 722 aagaagagagaggggag 739  
|||||  
Db 159771 AAGAAAGAGAGGGGAG 159754

RESULT 7  
US-08-724-394A-21/c  
Sequence 21, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONTIG"

TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONTIG"  
US-08-724-394A-21

Query Match 1.9%; Score 18; DB 3; Length 246240;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 722 aagaagagagaggggag 739  
|||||  
Db 159771 AAGAAAGAGAGGGGAG 159754

RESULT 8  
US-08-724-394A-22/c  
Sequence 22, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Kronmal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONTIG"



US-08-724-394A-22

Query Match 1.9%; Score 18; DB 3; Length 246240;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 722 aagaagagagagagag 739  
|||||

Db 159771 AAGAAAGAGAGAGGAG 159754

## RESULT 9

US-08-465-794-4

; Sequence 4, Application US/08465794

; Patent No. 5886141

; GENERAL INFORMATION:

; APPLICANT: FOLKMAN, MOSES J.

; APPLICANT: SHING, YUEN

; APPLICANT: IGARASHI, KOICHI

; TITLE OF INVENTION: SMOOTH MUSCLE MITOGEN AND ISOLATED DNA

; TITLE OF INVENTION: CODING THEREFORE

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &amp;

; ADDRESSEE: CUSHMAN

; STREET: 130 WATER STREET

; CITY: BOSTON

; STATE: MASSACHUSETTS

; COUNTRY: US

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/465,794

; FILING DATE:

; CLASSIFICATION: 530

; APPLICATION DATA:

; APPLICATION NUMBER: US 08/007,126

; FILING DATE: 21-JAN-1993

; APPLICATION NUMBER: US 07/994,776

; FILING DATE: 22-DEC-1992

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/872,597

; FILING DATE: 23-APR-1992

; APPLICATION NUMBER: US 07/872,792

; FILING DATE: 23-APR-1992

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/833,552

; FILING DATE: 10-FEB-1992

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/832,939

; FILING DATE: 10-FEB-1992

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/766,354

; FILING DATE: 26-SEP-1991

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/604,778

; FILING DATE: 26-OCT-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: RESNICK, DAVID S.

; REGISTRATION NUMBER: 34235

; REFERENCE/DOCKET NUMBER: 40435-CIP-8

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 523-3400

; TELEFAX: (617) 523-6440

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

LENGTH: 1175 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Mouse  
STRAIN: BTC-JC10  
CELL TYPE: Tumor Cell  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 113..643  
NAME/KEY: met peptide  
LOCATION: 206..643  
US-08-465-794-4

Query Match 1.8%; Score 17; DB 3; Length 1179;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 aaccacacacacacac 959  
|||||

Db 1119 AACCAACCAACCAACCA 1135

## RESULT 10

US-09-049-813-4

; Sequence 4, Application US/09049813

; Patent No. 6013762

; GENERAL INFORMATION:

; APPLICANT: FOLKMAN, MOSES J.

; APPLICANT: SHING, YUEN

; APPLICANT: IGARASHI, KOICHI

; TITLE OF INVENTION: SMOOTH MUSCLE MITOGEN AND ISOLATED DNA

; TITLE OF INVENTION: CODING THEREFORE

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &amp;

; ADDRESSEE: CUSHMAN

; STREET: 130 WATER STREET

; CITY: BOSTON

; STATE: MASSACHUSETTS

; COUNTRY: US

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/049,813

; FILING DATE:

; CLASSIFICATION:

; APPLICATION DATA:

; APPLICATION NUMBER: 08/465,794

; FILING DATE:

; APPLICATION NUMBER: US 08/007,126

; FILING DATE: 21-JAN-1993

; APPLICATION NUMBER: US 07/994,776

; FILING DATE: 22-DEC-1992

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/872,597

; FILING DATE: 23-APR-1992

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/872,792

; FILING DATE: 23-APR-1992

; APPLICATION DATA:

; APPLICATION NUMBER: US 07/833,552

; FILING DATE: 10-FEB-1992

; APPLICATION DATA:

; APPLICATION DATA:

APPLICATION NUMBER: US 07/832,939  
FILING DATE: 10-FEB-1992  
PRIOR APPLICATION NUMBER: US 07/766,354  
FILING DATE: 26-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/604,778  
FILING DATE: 26-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, DAVID S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 40435-CIP-8  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 523-3400  
TELEFAX: (617) 523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1179 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Mouse  
STRAIN: BTC-JC10  
CELL TYPE: Tumor Cell  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 113..643  
NAME/KEY: mat peptide  
LOCATION: 206..643  
US-09-049-813-4

Query Match 1.8%; Score 17; DB 5; Length 1179;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 aaccaacacacacacac 959  
|||||

Db 1119 AACCAACCAACCAACCA 1135

RESULT 11  
US-08-633-879C-17/c  
Sequence 17, Application US/08633879C  
Patent No. 5928922  
GENERAL INFORMATION:  
APPLICANT: Kivirikko, Kari I.  
APPLICANT: Pihlajaniemi, Taina  
APPLICANT: Helakoski, Tarja I.  
APPLICANT: Annunen, Pia P.  
APPLICANT: Nissi, Ritva K.  
APPLICANT: No. 5928922elainen, Minna K.  
TITLE OF INVENTION: '2 SUBUNIT OF PROLYL-4-HYDROXYLASE  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING SUCH SUBUNIT AND  
TITLE OF INVENTION: METHODS FOR PRODUCING THE SAME  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds, LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2811  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,879C  
FILING DATE: 10-APR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Adams, Samuel B.  
REGISTRATION NUMBER: 30,605  
REFERENCE/DOCKET NUMBER: 8389-0041-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 630-493-4935  
TELEFAX: 650-493-5556  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 96..272  
OTHER INFORMATION:  
US-08-633-879C-17

Query Match 1.8%; Score 17; DB 4; Length 2121;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 gctgttcataactgtga 402  
|||||

Db 1013 GCTGTCATAACTGTGA 997

RESULT 12  
PCT-US92-00731-12/c  
Sequence 12, Application PC/TUS9200731  
GENERAL INFORMATION:  
APPLICANT: Kawabata, Shunichiro  
APPLICANT: Davie, Earl W.  
TITLE OF INVENTION: MICROSOMAL ENDOPEPTIDASE  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry  
STREET: 6300 Columbia Center  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/00731  
FILING DATE: 19920128  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/646,997  
FILING DATE: 28-JAN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 990008.415PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)622-4900  
TELEFAX: (206)682-6031  
TELEX: 3723836  
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 2148 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FRAGMENT TYPE: N-terminal  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 11..2075  
 PCT-US92-00731-12

Query Match 1.8%; Score 17; DB 6; Length 2148;

Best Local Similarity 100.0%; Pred. No. 25; Mismatches 0; Indels 0; Gaps 0;

QY 661 ttcttttttttcaaac 677  
 DB 1198 TTCCTTTTCTTCAAC 1182

RESULT 13

US-08-348-518C-3/c

Sequence 3, Application US/08348518C

Patent No. 6022740

GENERAL INFORMATION:

APPLICANT: SUDOL, MARIUS

APPLICANT: PEER, BORK

APPLICANT: HENRY, CHEN

TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A

TITLE OF INVENTION: SIGNALLING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE

TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES

TITLE OF INVENTION: THEREOF

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/348,518C

FILING DATE: 01-DEC-1994

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 515 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Human

IMMEDIATE SOURCE:

CLONE: YAP

FEATURE:

NAME/KEY: CDS  
 LOCATION: 275..1637  
 US-08-348-518C-3

Query Match 1.8%; Score 17; DB 5; Length 515;

Best Local Similarity 100.0%; Pred. No. 24;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 943 acccaaccaaccaacca 959  
 DB 2415 ACCCAACCAACCAACCA 2399

RESULT 14

US-08-476-509B-3/c

Sequence 3, Application US/08476509B

Patent No. 6034212

GENERAL INFORMATION:

APPLICANT: SUDOL, MARIUS

APPLICANT: PEER, BORK

APPLICANT: HENRY, CHEN

TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A

TITLE OF INVENTION: SIGNALLING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE

TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES

TITLE OF INVENTION: THEREOF

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/476,509B

FILING DATE: 01-DEC-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-101 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 515 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Human

IMMEDIATE SOURCE:

CLONE: YAP

FEATURE:

NAME/KEY: CDS

LOCATION: 275..1637

US-08-476-509B-3

Query Match 1.8%; Score 17; DB 5; Length 515;

Best Local Similarity 100.0%; Pred. No. 24;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 01:08:50 ; Search time 3299.68 Seconds  
(without alignments)  
1798.809 Million cell updates/sec

Title: US-09-101-423A-5  
Perfect score: 960  
Sequence: 1 gagggggtgtgtgacagtt.....tcaaccacccaacacacat 960

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 7189864 seqs, 3091403243 residues

Word size : 0

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:

- 1: gb\_est1.\*
- 2: gb\_est2.\*
- 3: gb\_est3.\*
- 4: gb\_est4.\*
- 5: gb\_est5.\*
- 6: gb\_est6.\*
- 7: gb\_est7.\*
- 8: gb\_est8.\*
- 9: gb\_est9.\*
- 10: gb\_est10.\*
- 11: gb\_est11.\*
- 12: gb\_est12.\*
- 13: gb\_est13.\*
- 14: gb\_est14.\*
- 15: gb\_est15.\*
- 16: gb\_est16.\*
- 17: gb\_est17.\*
- 18: gb\_est18.\*
- 19: gb\_est19.\*
- 20: gb\_est20.\*
- 21: gb\_est21.\*
- 22: gb\_est22.\*
- 23: gb\_est23.\*
- 24: gb\_est24.\*
- 25: gb\_est25.\*
- 26: gb\_est26.\*
- 27: gb\_est27.\*
- 28: gb\_est28.\*
- 29: gb\_est29.\*
- 30: gb\_est30.\*
- 31: gb\_est31.\*
- 32: gb\_est32.\*
- 33: gb\_est33.\*
- 34: gb\_est34.\*
- 35: gb\_est35.\*
- 36: gb\_est36.\*
- 37: gb\_est37.\*
- 38: gb\_est38.\*
- 39: gb\_est39.\*
- 40: gb\_est40.\*
- 41: em\_estba.\*
- 42: em\_estfun.\*
- 43: em\_esthum1.\*

- 44: em\_esthum2.\*
- 45: em\_esthum3.\*
- 46: em\_esthum4.\*
- 47: em\_esthum5.\*
- 48: em\_esthum6.\*
- 49: em\_esthum7.\*
- 50: em\_esthum8.\*
- 51: em\_esthum9.\*
- 52: em\_esthum10.\*
- 53: em\_esthum11.\*
- 54: em\_esthum12.\*
- 55: em\_esthum13.\*
- 56: em\_esthum14.\*
- 57: em\_esthum15.\*
- 58: em\_esthum16.\*
- 59: em\_esthum17.\*
- 60: em\_esthum18.\*
- 61: em\_esthum19.\*
- 62: em\_esthum20.\*
- 63: em\_estin1.\*
- 64: em\_estin2.\*
- 65: em\_estin3.\*
- 66: em\_estin4.\*
- 67: em\_estom.\*
- 68: em\_estov1.\*
- 69: em\_estov2.\*
- 70: em\_estp11.\*
- 71: em\_estp12.\*
- 72: em\_estp13.\*
- 73: em\_estp14.\*
- 74: em\_estp15.\*
- 75: em\_estro1.\*
- 76: em\_estro2.\*
- 77: em\_estro3.\*
- 78: em\_estro4.\*
- 79: em\_estro5.\*
- 80: em\_estro6.\*
- 81: em\_estro7.\*
- 82: em\_estro8.\*
- 83: em\_estro9.\*
- 84: em\_estro10.\*
- 85: em\_estro11.\*
- 86: em\_estro12.\*
- 87: em\_estro13.\*
- 88: gb\_gss1.\*
- 89: gb\_gss2.\*
- 90: gb\_gss3.\*
- 91: gb\_gss4.\*
- 92: em\_gss1.\*
- 93: em\_gss2.\*
- 94: em\_gss3.\*
- 95: em\_gss4.\*
- 96: gb\_gss5.\*
- 97: gb\_gss6.\*
- 98: gb\_gss7.\*
- 99: gb\_gss8.\*
- 100: gb\_gss9.\*
- 101: em\_gss5.\*
- 102: em\_gss6.\*
- 103: em\_gss7.\*
- 104: em\_gss8.\*
- 105: em\_gss9.\*
- 106: em\_gss10.\*
- 107: em\_gss11.\*
- 108: gb\_gss10.\*
- 109: gb\_gss11.\*
- 110: em\_gss12.\*
- 111: gb\_gss12.\*
- 112: gb\_gss13.\*
- 113: gb\_gss14.\*
- 114: gb\_gss15.\*
- 115: gb\_gss16.\*
- 116: gb\_gss17.\*

117: gb-gss18:\*

118: gb-gss19:\*

119: em-gss13:\*

120: gb-gss20:\*

121: gb-gss21:\*

122: gb-gss22:\*

123: gb-gss23:\*

124: gb-gss24:\*

125: em-gss14:\*

126: em-gss15:\*

127: em-gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	36	3.8	369	5	AA619625	AA619625 v154d02.r
C 2	36	3.8	418	11	AI615880	AI615880 v154d02.y
C 3	30	3.1	619	115	AZ096005	AZ096005 RPCI-23-4
C 4	27	2.8	393	6	AA850795	AA850795 EST193563
C 5	27	2.8	411	7	AA875453	AA875453 UI-R-E0-C
C 6	27	2.8	438	25	AW915147	AW915147 EST346451
C 7	26	2.7	127	114	AZ038972	AZ038972 RPCI-23-2
C 8	26	2.7	444	113	AQ976346	AQ976346 RPCI-23-3
C 9	26	2.7	483	6	AA709784	AA709784 vv41901.r
C 10	26	2.7	570	113	AQ979459	AQ979459 RPCI-23-2
C 11	26	2.7	639	115	AZ109019	AZ109019 RPCI-23-4
C 12	26	2.7	697	115	AZ123055	AZ123055 RPCI-23-3
C 13	25	2.6	231	118	AZ300070	AZ300070 RPCI-23-1
C 14	25	2.6	424	114	AZ079194	AZ079194 RPCI-23-4
C 15	25	2.6	433	11	AI593877	AI593877 v125b10.y
C 16	24	2.5	128	117	AZ274962	AZ274962 RPCI-23-1
C 17	24	2.5	324	115	AZ116348	AZ116348 RPCI-23-1
C 18	24	2.5	453	117	AZ274879	AZ274879 RPCI-23-1
C 19	24	2.5	497	25	AW910590	AW910590 ur85d12.x
C 20	24	2.5	689	23	AW557054	AW557054 L0276G05-
C 21	24	2.5	727	10	AI437387	AI437387 md04d05.y
C 22	24	2.5	728	23	AW557056	AW557056 L0276G07-
C 23	23	2.4	267	113	AQ980090	AQ980090 RPCI-23-3
C 24	23	2.4	274	33	BE100871	BE100871 UI-R-BJ1-
C 25	23	2.4	313	21	AW252793	AW252793 UI-R-BJ0-
C 26	23	2.4	319	12	AI707038	AI707038 UI-R-AD1-
C 27	23	2.4	337	117	AZ272747	AZ272747 RPCI-23-8
C 28	23	2.4	373	114	AQ030326	AQ030326 RPCI-23-3
C 29	23	2.4	384	21	AW253796	AW253796 UI-R-BJ0-
C 30	23	2.4	473	21	AW252679	AW252679 UI-R-BJ0-
C 31	23	2.4	485	7	AI010317	AI010317 EST204768
C 32	23	2.4	493	117	AZ251034	AZ251034 RPCI-23-4
C 33	23	2.4	563	115	AZ118429	AZ118429 RPCI-23-4
C 34	23	2.4	564	115	AI100086	AI100086 RPCI-23-4
C 35	22	2.3	243	22	AW344425	AW344425 UI-R-BJ0p
C 36	22	2.3	296	15	AV045634	AV045634 AV045634
C 37	22	2.3	302	7	AA996587	AA996587 UI-R-C0-h
C 38	22	2.3	355	117	AZ258644	AZ258644 RPCI-23-1
C 39	22	2.3	374	6	AA716790	AA716790 vu61e10.r
C 40	22	2.3	374	12	AI662327	AI662327 ms10e02.x
C 41	22	2.3	387	114	AZ055860	AZ055860 RPCI-23-4
C 42	22	2.3	435	113	AQ980833	AQ980833 RPCI-23-2
C 43	22	2.3	455	19	AW048192	AW048192 UI-M-BH1-
C 44	22	2.3	505	19	AW060679	AW060679 UI-M-BH1-
C 45	22	2.3	522	113	AQ997454	AQ997454 RPCI-23-3

## ALIGNMENTS

RESULT 1  
AA619625/c

LOCUS  
DEFINITION  
AA619625 369 bp mRNA EST 09-OCT-1997  
IMAGE:976035 5', mRNA sequence.  
ACCESSION  
AA619625  
VERSION  
AA619625:1 GI:2523501  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
REFERENCE  
1 (bases 1 to 369)  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kuabara, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
TITLE  
The WashU-HMI Mouse EST Project  
JOURNAL  
Unpublished (1996)  
COMMENT  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 358.  
Location/Qualifiers  
1..369  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:976035"  
/clone.lib="Stratagene mouse skin (#937313)"  
/sex="females"  
/tissue-type="whole skin"  
/dev-stage="11 weeks old"  
/note="Organ: skin; Vector: pBluescript SK-; Site.1: EcoRI  
Site.2: XhoI; Cloned unidirectionally. Primer: Oligo  
dt. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'  
adaptor sequence: 5' GAATCGGCACGAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

## FEATURES

## Source

## BASE COUNT

## ORIGIN

Query Match 3.8%; Score 36; DB 5; Length 369;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 aggggagctcaattctctggggcaagtgtgatcttt 864  
|||||

Db 283 AGGGGAGTCAGTCTCTTGGGGCAAGTTGATCTTT 248  
|||||

## RESULT 2

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

AI615880 418 bp mRNA EST 21-APR-1999  
V154D02.y1 Stratagene mouse skin (#937313) Mus musculus cDNA clone  
IMAGE:976035 5', mRNA sequence.  
AI615880  
AI615880.1 GI:4625047  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 418)

## AUTHORS

Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, I., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

## TITLE

The WashU-NCI Mouse EST Project 1999

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:556763

This read is a RESEQUENCE of a previously sequenced mouse clone  
correct orientation)

Seq primer: -40RP from Gibco  
High quality sequence stop: 417.

## FEATURES

source

Location/Qualifiers  
1. 418  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:976035"  
/clone\_lib="Stratagene mouse skin (#937313)"  
/sex="females"  
/tissue\_type="whole skin"  
/dev\_stage="11 weeks old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: skin; Vector: pBluescript SK-; Site\_1: EcoRI  
; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dt. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTCTTTTCTTTT 3'

BASE COUNT  
ORIGIN

116 a 74 c 104 g 123 t 1 others

## Query Match

Best Local Similarity 3.8%; Score 36; DB 11; Length 418;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 agggagtgcaagttctctggggcaagttgatcttt 864

|||||

DB 283 AGGGGAGTCAAGTCTCTGGGGCAAGTTGATCTTT 248

## RESULT 3

AZ096005/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1. (bases 1 to 619)

Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akimret

, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other\_GSSs: RPCI-23-461114.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: zhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html

Plate: 461 row: 1 column: 14

Seq primer: T7

Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1. 619  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-461114"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT  
ORIGIN

203 a 117 c 101 g 198 t

## Query Match

Best Local Similarity 3.1%; Score 30; DB 115; Length 619;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 atttgatactgttatgaattgtaataaa 431

|||||

DB 510 ATTTCATCTGTATGAATTGTAATAAA 481

## RESULT 4

AA850795/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1. (bases 1 to 393)

Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,

Kerlavage, A.R. and Adams, M.D.

Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat

Gene Index

Unpublished (1998)

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13-21.

Location/Qualifiers

1. 393

/organism="Rattus sp."

/db\_xref="ATCC (inhost):2009196"

```

/db_xref="taxon:10118"
/clone="ROVAK85"
/clone_lib="Normalized rat ovary, Bento Soares"
/note="Organ: ovary; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT      114 a      65 c      120 g      94 t
ORIGIN

Query Match      2.8%; Score 27; DB 6; Length 393;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 ttgaccctttaatacagtcgctcatg 348
|||||
Db 373 TTGACCCCTTAAATACAGTGCCTCATG 347
|||||

RESULT 5
AA875453/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

AA875453      411 bp      mRNA      EST      04-JUL-1999
UI-R-E0-cp-b-06-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-cp-b-06-0-UI.3, similar to gi1279555|emb|X80638|MMRHOCR
M.musculus rhoc mRNA, mRNA sequence.
AA875453
AA875453.1 GI:4231331
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 411)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
On Mar 20, 1998 this sequence version replaced gi:2980401.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dt track served to identify it as a clone from the normalized
adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
ID-1777675 The following repetitive elements were found in this
cDNA sequence: 17-111, >URR1B#DNA/MER1_type
Seq primer: M13 Forward
POLYA-No.
FEATURES
source
Location/Qualifiers
1..411
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-cp-b-06-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DHI0B (Life Technologies)"
/polylinker="Vector: pT73D-Pac (Pharmacia) with a modified
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dt track which allows
identification of the library of origin of a clone within
the mixture."

```

```

BASE COUNT      96 a      78 c      130 g      107 t
ORIGIN

Query Match      2.8%; Score 27; DB 7; Length 411;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 ttgaccctttaatacagtcgctcatg 348
|||||
Db 92 TTGACCCCTTAAATACAGTGCCTCATG 66
|||||

RESULT 6
AW915147
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AW915147      438 bp      mRNA      EST      25-MAY-2000
EST346451 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
RGICL10.3, end, mRNA sequence.
AW915147
AW915147.1 GI:8080834
Rattus sp.
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 438)
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavags,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information
Seq primer: M13 Reverse.
FEATURES
source
Location/Qualifiers
1..438
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="RGICL10"
/clone_lib="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      105 a      131 c      74 g      128 t
ORIGIN

Query Match      2.8%; Score 27; DB 25; Length 438;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 ttgaccctttaatacagtcgctcatg 348
|||||
Db 31 TTGACCCCTTAAATACAGTGCCTCATG 57
|||||

RESULT 7
AZ038972
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AZ038972      127 bp      DNA      GSS      01-MAR-2000
RPCI-23-242K22.t1v RPCI-23 Mus musculus genomic clone RPCI-23-242K22
, DNA sequence.
AZ038972
AZ038972.1 GI:7128969
GSS.
house mouse.
Mus musculus

```



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 127)

REFERENCE  
AUTHORS  
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.

TITLE  
JOURNAL  
COMMENT  
Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 242 row: K column: 22  
Seq primer: T7  
Class: BAC ends.

#### FEATURES

Location/Qualifiers  
1..127  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-242K22"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 34 a 27 c 20 g 46 t  
ORIGIN

Query Match 2.7%; Score 26; DB 114; Length 127;  
Best Local Similarity 100.0%; Pred. No. 0.019; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0;

QY 409 tactgttatgaattgtaataataa 434  
|||||  
DB 96 TACTGTATGATTGTAATAATAA 121

RESULT 8  
AAQ976346/c 444 bp DNA GSS 29-JAN-2000  
LOCUS  
DEFINITION  
RPCI-23-335017.TJ RPCI-23 Mus musculus genomic clone RPCI-23-335017  
DNA sequence.  
ACCESSION  
AQ976346  
VERSION  
AQ976346.1 GI:6808658  
KEYWORDS  
GSS.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 444)

REFERENCE  
AUTHORS  
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.

TITLE  
JOURNAL  
COMMENT  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics

The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 335 row: O column: 17  
Seq primer: SP6  
Class: BAC ends.

#### FEATURES

Location/Qualifiers  
1..444  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-335017"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 136 a 77 c 106 g 125 t  
ORIGIN

Query Match 2.7%; Score 26; DB 113; Length 444;  
Best Local Similarity 100.0%; Pred. No. 0.019; Indels 0; Gaps 0;  
Matches 26; Conservative 0; Mismatches 0;

QY 835 gtcaagttctctggcgcaagttgat 860  
|||||  
DB 163 GTCAGTTCCTTGGGCGCAAGTTGAT 138

#### RESULT 9

AA709784/c 483 bp mRNA EST 24-DEC-1997  
LOCUS  
DEFINITION  
v41g01.r1 Soares\_thymus\_2NBMT Mus musculus cDNA clone  
IMAGE:1225008 5', mRNA sequence.

ACCESSION  
AA709784  
VERSION  
AA709784.1 GI:2719702  
KEYWORDS  
EST.  
SOURCE  
house mouse.

ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 483)

REFERENCE  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE  
JOURNAL  
COMMENT  
The WashU-HMI Mouse EST Project  
Unpublished (1996)

Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:650600



with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT 192 a 136 c 163 g 148 t

ORIGIN

Query Match 2.7%; Score 26; DB 115; Length 639;

Best Local Similarity 100.0%; Pred. No. 0.02; Mismatches 0; Indels 0; Gaps 0;

Matches 26; Conservative 0;

QY 409 tactgttatgaattgataataaata 434

DB 436 TACTGTTATGAGTCTGTAATATAATA 411

RESULT 12

AZ123055

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other GSSs: RPCI-23-30F22-TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html

Plate: 30 row: F column: 22

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. .697

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-30F22"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 214 a 135 c 130 g 218 t

ORIGIN

Query Match

2.7%; Score 26; DB 115; Length 697;

Best Local Similarity 100.0%; Pred. No. 0.02; Mismatches 26; Conservative 0; Indels 0; Gaps 0;

QY 409 tactgttatgaattgataataaata 434

DB 193 TACTGTTATGAGTCTGTAATATAATA 218

RESULT 13

AZ300070/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other GSSs: RPCI-23-120A1-TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html

Plate: 120 row: A column: 1

Seq primer: T7

Class: FAC ends.

Location/Qualifiers

1. .231

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-120A1"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 42 a 61 c 54 g 74 t

ORIGIN

Query Match 2.6%; Score 25; DB 118; Length 231;

Best Local Similarity 100.0%; Pred. No. 0.062;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 839 agttcccttgaggcaagtttgatctt 863

DB 145 AGTTCCTTGGGCAAGTTTGATCTT 121

RESULT 14

```

AZ079194/c
LOCUS      424 bp      DNA      31-MAR-2000
DEFINITION RPCI-23-438G4-TV RPCI-23 Mus musculus genomic clone RPCI-23-438G4,
            DNA sequence.
ACCESSION  AZ079194
VERSION     AZ079194.1 GI:7372093
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 424)
AUTHORS   Zhao,S., Niekman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levings,M., McGann,S., Tsagaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-23
            Unpublished (1999)
            Other_GSSs: RPCI-23-438G4.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buhalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buhalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 438 row: G column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..424
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-438G4"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1:
                     ECORI; Site:2: ECORI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase. Size
                     selected DNA was cloned into the pBACE3.6 vector at the
                     EcoRI sites. The ligation products were transformed into
                     DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT          137 a      93 c      89 g      105 t
ORIGIN
Query Match      2.6%; Score 25; DB 114; Length 424;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 410 actgttatgaattgtaataataaata 434
|||||
Db 384 ACTGTTATGAATTGCTAATAATAAATA 360

RESULT 15
AI593877/c
LOCUS      433 bp      mRNA      21-APR-1999
DEFINITION vt25b10.y1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1121659 5', mRNA sequence.
ACCESSION  AI593877
VERSION     AI593877.1 GI:4602925
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 433)
AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Watson,R. and Wilson,R.
            The WashU-NCI Mouse EST Project 1999
            Unpublished (1999)
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LINL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:610995
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 360.
FEATURES             Location/Qualifiers
     source           1..433
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1121659"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /note="Vector: p7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5',
                     TGTTACGAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified p7T3 vector.
                     Library constructed by Bob Barstead. The C2C12 cell line
                     (available from ATCC, catalog # CRL-1772) differentiates
                     rapidly, forming contractile myotubes and producing
                     characteristic muscle proteins."
BASE COUNT          115 a      73 c      119 g      126 t
ORIGIN
Query Match      2.6%; Score 25; DB 11; Length 433;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 839 agttcttggggcaagttgatctt 863
|||||
Db 292 AGTCTCTGGGCGCAAGTTGATCTT 268

Search completed: November 4, 2000, 01:08:54
Job time: 17435 sec

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 01:08:29 ; Search time 7109.34 Seconds  
(without alignments)  
589.770 Million cell updates/sec

Title: US-09-101-423A-5  
Perfect score: 960  
Sequence: 1 gagggggtggtgcacagtt.....tcaaccaaccaaccacat 960

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_bal:.\*  
2: gb\_ba2:.\*  
3: gb\_om:.\*  
4: gb\_ov:.\*  
5: gb\_pat:.\*  
6: gb\_ph:.\*  
7: gb\_pl1:.\*  
8: gb\_pl2:.\*  
9: gb\_pr1:.\*  
10: gb\_pr2:.\*  
11: gb\_pr3:.\*  
12: em\_fun:.\*  
13: em\_hum1:.\*  
14: em\_hum2:.\*  
15: em\_in:.\*  
16: em\_om:.\*  
17: em\_or:.\*  
18: em\_ov:.\*  
19: em\_pat:.\*  
20: em\_ph:.\*  
21: em\_pl:.\*  
22: em\_ro:.\*  
23: em\_sts:.\*  
24: em\_sy:.\*  
25: em\_un:.\*  
26: em\_vi:.\*  
27: gb\_htgl:.\*  
28: gb\_htg2:.\*  
29: gb\_in1:.\*  
30: gb\_in2:.\*  
31: em\_bal:.\*  
32: em\_ba2:.\*  
33: em\_hum3:.\*  
34: em\_hum4:.\*  
35: gb\_pr4:.\*  
36: gb\_htg3:.\*  
37: gb\_htg4:.\*  
38: gb\_htg5:.\*  
39: gb\_htg6:.\*  
40: gb\_htg7:.\*  
41: em\_htgl:.\*  
42: em\_htg2:.\*  
43: em\_htg3:.\*

44: em\_hum5:.\*  
45: gb\_pl3:.\*  
46: gb\_pf5:.\*  
47: gb\_htg8:.\*  
48: gb\_htg9:.\*  
49: gb\_htg10:.\*  
50: gb\_htg11:.\*  
51: gb\_htg12:.\*  
52: gb\_htg13:.\*  
53: gb\_htg14:.\*  
54: gb\_in3:.\*  
55: gb\_htg15:.\*  
56: gb\_htg16:.\*  
57: gb\_htg17:.\*  
58: em\_htg4:.\*  
59: em\_htg5:.\*  
60: em\_htg6:.\*  
61: em\_htg7:.\*  
62: em\_hum6:.\*  
63: gb\_htg18:.\*  
64: gb\_htg19:.\*  
65: gb\_ba3:.\*  
66: em\_htg8:.\*  
67: em\_htg9:.\*  
68: em\_htg10:.\*  
69: em\_htg11:.\*  
70: em\_htg12:.\*  
71: em\_htg13:.\*  
72: em\_htg14:.\*  
73: em\_htg15:.\*  
74: em\_htg16:.\*  
75: em\_htg17:.\*  
76: em\_htg18:.\*  
77: em\_htg19:.\*  
78: em\_htg20:.\*  
79: em\_htg21:.\*  
80: em\_htg22:.\*  
81: em\_htg23:.\*  
82: gb\_pr6:.\*  
83: gb\_pr7:.\*  
84: gb\_htg20:.\*  
85: gb\_htg21:.\*  
86: gb\_htg22:.\*  
87: gb\_htg23:.\*  
88: gb\_ro:.\*  
89: gb\_sts1:.\*  
90: gb\_sts2:.\*  
91: gb\_sy:.\*  
92: gb\_un:.\*  
93: gb\_vil:.\*  
94: gb\_vil2:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	960	100.0	960	5	A64062
c 2	106.8	11.1	235586	49	AC024173 Mus muscu
c 3	106.8	11.1	269908	38	AC015535 Mus muscu
4	105.4	11.0	94444	56	AC069445 Mus muscu
5	103	10.7	237143	57	AC073769 Mus muscu
6	96.8	10.1	193432	57	AC073290 Mus muscu
c 7	96	10.0	2830	5	AR016448
c 8	96	10.0	2830	88	MMU43673
9	96	10.0	314146	57	AC073759 Mus muscu
10	95.2	9.9	178016	39	AC016087 Homo sapi
c 11	94.6	9.9	314146	57	AC073759 Mus muscu
12	94	9.8	191276	55	AC068140 Mus muscu

13 93.4 9.7 195190 53 AC044778  
 C 14 93.2 9.7 207160 39 AC016464  
 C 15 93.2 9.7 214616 37 AC074042  
 C 16 93 9.7 139171 50 AC024883  
 17 93 9.7 196889 57 AC073730  
 18 92.8 9.7 5366 88 AF22603853  
 19 92.8 9.7 132104 57 AC073672  
 20 92.2 9.6 204618 88 AF146793  
 21 92 9.6 110000 47 AC020850\_3  
 22 91.8 9.6 189105 27 AC007641  
 C 23 91.6 9.5 236554 48 AC022232  
 C 24 91.4 9.5 183156 48 AC023062  
 25 91.4 9.5 222119 57 AC073297  
 C 26 91.2 9.5 130049 57 AC073296  
 C 27 91.2 9.5 138258 57 AC074337  
 C 28 91.2 9.5 237143 57 AC073769  
 C 29 91 9.5 200275 57 AC073883  
 C 30 90 9.4 132104 57 AC073672  
 31 88.8 9.2 61245 56 AC069449  
 32 88.8 9.2 108754 40 AC018461  
 33 88.6 9.2 239574 57 AC073791  
 C 34 88.4 9.2 139258 57 AC073673  
 C 35 88.4 9.2 176245 38 AC018017  
 C 36 88.4 9.2 196146 49 AC023609  
 C 37 88 9.2 116102 27 AC073707  
 C 38 88 9.2 210861 57 AC073734  
 C 39 87.8 9.1 176633 56 AC069074  
 C 40 87.6 9.1 194515 85 AL365329  
 C 41 87.6 9.1 249404 85 AL365322  
 C 42 87.4 9.1 175371 57 AC073748  
 C 43 87.2 9.1 23595 88 MUSXPDG1  
 C 44 87.2 9.1 166513 57 AC073787  
 C 45 87.2 9.1 208980 53 AC053470

## ALIGNMENTS

RESULT 1  
 A64062  
 LOCUS A64062 960 bp DNA PAT 29-MAR-1999  
 DEFINITION Sequence 5 from Patent WO9725443.  
 ACCESSION A64062  
 VERSION A64062.1 GI:3717536  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unclassified.  
 REFERENCE 1 (bases 1 to 960)  
 AUTHORS Rudland, P.S. and Barraclough, B.R.  
 TITLE METASTASIS INDUCING DNA'S  
 JOURNAL Patent: WO 9725443-A 17-JUL-1997;  
 UNIV LIVERPOOL (GB)  
 FEATURES  
 Location/Qualifiers  
 1..960  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 256 a 193 c 233 g 278 t  
 ORIGIN

Query Match 100.0%; Score 960; DB 5; Length 960;  
 Best Local Similarity 100.0%; Pred. NO. 6.7e-237;  
 Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gagggtggtggcagcttatgttttttaggaagggttccatgaacctcagcaagct 60  
 DB 1 GAGGGGGGGGGCAGCTTATGTTTGTAGGAAGGGTCCATGAACCTCAGCAGACT 60  
 QY 61 cgggttagaatttaaacgctgagggaattttttttaaatcgctatgaattcgac 120  
 DB 61 CGGGTTAGAAATTTAAAGCCCTGAGGGGAATTTTTTTTAAATCGCTATGAATCTGAC 120

QY 121 atgagaaaaacagatcagaaacgttcttctgcttcagaaagacaaagtgtgtgagtaa 180  
 DB 121 ATGAGAAAAACAGATCAGAAACGTTCTTGTGCTTCAGAAAAGCACAAGTGTGTGAGCTAA 180  
 QY 181 cagactcacactgggttctgagagcacatctgataccagagagcgtcagataatgtcccc 240  
 DB 181 CAGACTGCACACTGCTGTTTCGAGGCACATCTGGATCAGAGGCGTCAGATAATGTCCCC 240  
 QY 241 aaagtgaaatgcattgtctgacagaccagtggtggtgggggtgctcacagccagc 300  
 DB 241 AAAGTAAATGCATTGCTTGCACAGTACCGAGTGTGGTGGGGGTGCCTACAGCCAGC 300  
 QY 301 ggttctcaactctcgtgactctcgaccctttaaatacagtcctcatgtctctgtgacct 360  
 DB 301 GGTTCCTCAACCTTCTCGATGCTTCGACCCCTTAAATACAGTGCCTCATGCTCTGGTACCT 360  
 QY 361 ccccaaccttaaatatttttctgctgttcataactgtgatttgcatactgttatgaa 420  
 DB 361 CCCCAACCTTAAATATTATTTTGTGCTGTTCATACTGTGATTTGTACTACTGTATGAA 420  
 QY 421 ttgtaataataataatttttgaagaagaggtttgccaaggggttggagaactgctgtcta 480  
 DB 421 TTGTAATATAATAATTTTGAAGAAAGAGGTTGCCAAGGGTTTGAGAACTGCTGTCTTA 480  
 QY 481 gcccacgtgaggttttttctgctcatttgggggtttttatgagcgagagcttatagcc 540  
 DB 481 GCCCAGCTGGATGGTGTTCGTCATTTGGGGTTTTTATGAGCGAGAGTCTTATGTAGCC 540  
 QY 541 caggctagcagcgtagaatgtctacttagctgaggaataaaccttggaactctctgagac 600  
 DB 541 CAGGCTAGCAGCGCTAGAAATGTGCTACTTAGCTGAGGAATAAACCTTGGAACCTTCTGAGGAC 600  
 QY 601 tgagagactggtttagtctcctaagaactggaataatagctgaggttggctactgtggg 660  
 DB 601 TGGAGAGACTGGCTTAGTCTCCTCAAGAAACTGGAAATAGCTGGAGTTGGCTACTTGTGGG 660  
 QY 661 ttcttttctcaaaccttttctactcttttccacctgtcgcccccttaacactaaa 720  
 DB 661 TTCTTTTCTTCAAACTTTTCTACTCTTTTCCACCTGTGGCCCCCTAACACTAAA 720  
 QY 721 taagaagagaaaggggagcatagaggggaaagaaacccctgaataacgtcagtagttg 780  
 DB 721 TAAGAAAGAGAAAGGGGAGCATAGAGGGGAAAGAAACCCCTGAATAACGTCAGTAGTTG 780  
 QY 781 gaaaaggggggtgacataatgttgcattagacacacatctcgtgattaggggagtcag 840  
 DB 781 GCAAAGGGGGGTGACATATGTTGTATTAGACCACATCCTGCTGATTAGGGGGAGTCAAG 840  
 QY 841 ttcttgggggcaagtttgatcttctggttaacgatatctaatcttctcctctgtgctt 900  
 DB 841 TTCTTTGGGGCAAGTTTGATCTTTCGTGTAAGATATCTAATTTCTTCCCTGTGCTT 900  
 QY 901 cgtcttggtaaacagacttgatacccaaatggaccatacaaccaaccaaccaaccat 960  
 DB 901 CGTCTTTGTGTACACGACTTGATAACCACCAATGGACCATCAACCAACCAACCAACCAT 960

RESULT 2  
 AC024173/c  
 LOCUS AC024173/c  
 DEFINITION  
 ACCESSION AC024173  
 VERSION AC024173.1 GI:7105519  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 235586)  
 Beckstrom-Sternberg, S.M., Benjamin, B., Bouffard, G.G.,  
 Dietrich, N.L., Eagle, W.O., Gan, W., Gupta, J., Ho, S.-L., Huang, M.C.,  
 Idol, J., Jamison, D.C., Lee-Lin, S.-Q., Maduro, Q.L., Maduro, V.B.,



Mastrian, D., McCloskey, J.C., Morse, E., Ojodu, M.A., Pearson, R.,  
Stantrop, S., Summers, J.O., Thomas, J.W., Thomas, P.J.,  
Tiongson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.D., Walker, M.A.,  
Wetherby, K.D., and Green, E.D.  
NISC Mouse Sequencing Initiative

Unpublished

2 (bases 1 to 235586)

TITLE  
JOURNAL  
AUTHORS  
JOURNAL

Direct Submission  
Submitted (25-FEB-2000) NIH Intramural Sequencing Center, 8717  
GroveMont Circle, Gaithersburg, MD 20877, USA

COMMENT

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: [nisc.mouse@nih.gov](mailto:nisc.mouse@nih.gov)

----- Project Information

Center project name: xd

Center clone name: 059P17

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 225818 bases at least Q40

Consensus quality: 229346 bases at least Q30

Consensus quality: 231308 bases at least Q20

Insert size: 211000; agarose-fp

Insert size: 275000; pulse-field-gel

Quality coverage: 6.48x in Q20 bases; agarose-fp

Quality coverage: 4.98x in Q20 bases; pulse-field-gel

Quality coverage: 5.81x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2331: contig of 2331 bp in length  
\* gap of unknown length  
\* 2332 contig of 4324 bp in length  
\* 6656 gap of unknown length  
\* 14028 contig of 7372 bp in length  
\* 20172 contig of 6145 bp in length  
\* 28247 contig of 8075 bp in length  
\* 39419 gap of unknown length  
\* 52271 gap of unknown length  
\* 69465 contig of 17194 bp in length  
\* 90891 gap of unknown length  
\* 114318 gap of unknown length  
\* 136987 contig of 22669 bp in length  
\* 154399 gap of unknown length  
\* 190170 contig of 17411 bp in length  
\* 235586 gap of unknown length  
\* contig of 35771 bp in length  
\* contig of 45417 bp in length.

FEATURES  
source

Location/Qualifiers  
1. .235586  
/organism="Mus musculus"  
/strain="C57BL6/J"  
/db\_xref="taxon:10090"

/chromosome="13"

/clone="RP23-59P17"

/clone\_lib="RBCI mouse BAC library 23"

BASE COUNT 69916 a 48971 c 48532 g 88134 t 33 others

ORIGIN

Query Match 11.1%; Score 106.8; DB 49; Length 235586;  
Best Local Similarity 76.9%; Pred No. 5.5e-17;  
Matches 143; Conservative 0; Mismatches 42; Indels 1; Gaps 1;  
QY 297 cagcgggtctcaacctctcgtatgcttcgaccttataatcagtgctcctcgtcgtggtg 356  
DB 93532 CAGTGTCTCGAACCTTCTTAATGCTGAGACCTTTAATACAGTCCATATGTTGTGATG 93473  
QY 357 acctcccaaccttaaaattattttgttgcct-gttcataaactggtgatttatactgtt 415  
DB 93472 ACCCCCCACCATAAATATTTTGTGCTACTTCACTAACTGTAATGCTACTGTT 93413  
QY 416 atgaattgtaataataattttgagaaaggggttcccaaggggttgagaaactgctg 475  
DB 93412 ATGAATCTTATATAAACTTTCAGAGATAGAGGTGCTGCTAAAGGGTTGGACACACAG 93353  
QY 476 ttctag 481  
DB 93352 GTTGAG 93347

RESULT 3

AC015535/c

LOCUS

DEFINITION

AC015535 269908 bp DNA HTG 14-MAR-2000

Mus musculus clone RP23-39M18, WORKING DRAFT SEQUENCE, 20 unordered

pieces.

AC015535

VERSION

AC015535.4 GI:7230304

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

KEYWORDS

house mouse

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 269908)

Birren, E., Linton, L., Nusbaum, C. and Lander, E.

Mus musculus, clone RP23-39M18

Unpublished

2 (bases 1 to 269908)

Birren, E., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Castle, A., Collins, S., Collins, S., Collymore, A.,

Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wynan, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 12, 2000 this sequence version replaced gi:6504498.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L909



Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,  
 Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,  
 Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,  
 Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,  
 Gorrell, J. H., Gunaratne, P., Haller, G., Hernandez, J., Hogue, M.,  
 Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,  
 Kovar, C., Liu, J., Liu, W., Louissegh, H., Lozano, R. J., Martin, R.,  
 Massey, E., McLeod, M. P., Mei, G., Moore, S., Morgan, M., Morris, B.,  
 Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,  
 Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,  
 Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,  
 Muzny, D. M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,  
 Worley, K. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 94444)  
 Worley, K.C.  
 Direct Submission  
 Submitted (30-MAY-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: MAFA  
 Center clone name: RP23-270C13  
 ----- Summary Statistics  
 Sequencing vector: M13; L08821  
 Chemistry: Dye-Primer Bodypiz: 34% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 49562 bases at least Q40  
 Consensus quality: 73835 bases at least Q30  
 Consensus quality: 84393 bases at least Q20  
 Estimated insert size: 26253; sum-of-contigs estimation  
 Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 52 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2431: contig of 2431 bp in length  
 \* 2432 2531: gap of unknown length  
 \* 2532 4771: contig of 2240 bp in length  
 \* 4772 4871: gap of unknown length  
 \* 4872 7404: contig of 2533 bp in length  
 \* 7405 7505: gap of unknown length  
 \* 7506 10367: contig of 2863 bp in length  
 \* 10368 10468: gap of unknown length  
 \* 10469 12574: contig of 2106 bp in length  
 \* 12575 12673: gap of unknown length  
 \* 12674 15834: contig of 3161 bp in length  
 \* 15835 15934: gap of unknown length  
 \* 15935 17857: contig of 1923 bp in length  
 \* 17858 20991: gap of unknown length  
 \* 20992 21091: contig of 3034 bp in length  
 \* 21092 23910: gap of unknown length  
 \* 23911 24010: contig of 2819 bp in length  
 \* 24011 25942: gap of unknown length  
 \* 25943 26042: contig of 1932 bp in length  
 \* 26043 28158: gap of unknown length  
 \* 28159 30502: contig of 2116 bp in length  
 \* 30503 30602: gap of unknown length  
 \* 30603 32335: contig of 2244 bp in length  
 \* 32336 32335: contig of 1733 bp in length

\* 32336 32435: gap of unknown length  
 \* 32436 34387: contig of 1952 bp in length  
 \* 34388 34487: gap of unknown length  
 \* 34488 36071: contig of 1584 bp in length  
 \* 36072 36171: gap of unknown length  
 \* 36172 37872: contig of 1701 bp in length  
 \* 37873 37972: gap of unknown length  
 \* 37973 39893: contig of 1921 bp in length  
 \* 39894 39993: gap of unknown length  
 \* 39994 41854: contig of 1860 bp in length  
 \* 41855 43378: gap of unknown length  
 \* 43379 43478: contig of 1425 bp in length  
 \* 43479 45038: gap of unknown length  
 \* 45039 45138: contig of 1560 bp in length  
 \* 45139 47169: gap of unknown length  
 \* 47170 47269: contig of 2031 bp in length  
 \* 47270 49379: gap of unknown length  
 \* 49380 51029: contig of 2110 bp in length  
 \* 51030 51129: gap of unknown length  
 \* 51130 53321: contig of 1550 bp in length  
 \* 53322 53421: gap of unknown length  
 \* 53422 55052: contig of 2192 bp in length  
 \* 55053 55152: gap of unknown length  
 \* 55153 57018: contig of 1631 bp in length  
 \* 57019 57118: gap of unknown length  
 \* 57119 58660: contig of 1866 bp in length  
 \* 58661 58761: gap of unknown length  
 \* 58762 60074: contig of 1542 bp in length  
 \* 60075 61664: gap of unknown length  
 \* 61665 61764: contig of 1214 bp in length  
 \* 61765 63204: gap of unknown length  
 \* 63205 64872: contig of 1590 bp in length  
 \* 64873 64973: gap of unknown length  
 \* 64974 66128: contig of 1440 bp in length  
 \* 66129 66228: gap of unknown length  
 \* 66229 67550: contig of 1568 bp in length  
 \* 67551 67751: gap of unknown length  
 \* 67752 69847: contig of 1156 bp in length  
 \* 69848 69947: gap of unknown length  
 \* 69949 72045: contig of 1522 bp in length  
 \* 72046 72146: gap of unknown length  
 \* 72147 73753: contig of 1997 bp in length  
 \* 73754 73853: gap of unknown length  
 \* 73854 74954: contig of 2098 bp in length  
 \* 74955 75054: gap of unknown length  
 \* 75055 76141: contig of 1608 bp in length  
 \* 76142 76241: gap of unknown length  
 \* 76242 77352: contig of 1101 bp in length  
 \* 77353 77451: gap of unknown length  
 \* 77452 78633: contig of 1087 bp in length  
 \* 78634 78733: gap of unknown length  
 \* 78734 80165: contig of 1110 bp in length  
 \* 80166 80265: gap of unknown length  
 \* 80266 81756: contig of 1432 bp in length  
 \* 81757 81856: gap of unknown length  
 \* 81857 83119: contig of 1490 bp in length  
 \* 83120 83219: gap of unknown length  
 \* 83220 84591: contig of 1263 bp in length  
 \* 84592 84691: gap of unknown length  
 \* 84692 85853: contig of 1273 bp in length  
 \* 85854 85953: gap of unknown length  
 \* 85954 87208: contig of 1262 bp in length  
 \* 87209 87308: gap of unknown length  
 \* 87310 88317: contig of 1255 bp in length  
 \* 88318 88417: gap of unknown length  
 \* 88418 89424: contig of 1008 bp in length  
 \* 89425 89524: gap of unknown length  
 \* 89525 90660: contig of 1007 bp in length  
 \* 90661 90760: contig of 1137 bp in length  
 \* 90761 90760: gap of unknown length

\* 90761 91823: contig of 1063 bp in length  
 \* 91824 91923: gap of unknown length  
 \* 91924 93068: contig of 1145 bp in length  
 \* 93069 93168: gap of unknown length  
 \* 93169 94444: contig of 1276 bp in length.

BASE COUNT	24453 a	20485 c	19990 g	24354 t	5162 others
ORIGIN					

Query Match 11.0%; Score 105.4; DB 56; Length 94444;  
Best Local Similarity 83.2%; Pred. No. 1.12e-16;  
Matches 144; Conservative 0; Mismatches 26; Indels 3; Gaps 2;

- \* arbitrary. Gaps between the contigs are represented as
- \* runs of N, but the exact sizes of the gaps are unknown.
- \* This record will be updated with the finished sequence
- \* as soon as it is available and the accession number will
- \* be preserved.

BASE COUNT	ORIGIN
64366	clone_1110-RFC1 mouse BAC library 23"
53256	c
53088	g
64331	t
2102	others

```
Query Match      : 10.7%; Score 103; DB 57; Length 237143;
Best Local Similarity 74.6%; Pred. No. 5.2e-16;
Matches 156; Conservative 0; Mismatches 50; Indels 3; Gaps 2;
```

Qy 459 ggggttagaactgctgttcttagccac 487  
Db 49776 AGGTAATAACTCAAGGTTAAGAACCCAC 49804

RESULT 6  
AC073290 193432 bp DNA HTG 13-JUN-2000  
LOCUS Mus musculus clone RP23-441D12, WORKING DRAFT SEQUENCE, 29  
DEFINITION unordered pieces.  
AC073290 1 GI:8493565  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS mouse mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
1 (bases 1 to 193432)  
McCombie, W.R., Baker, J.P., Bahret, A., Bal, R., Dedhia, N.N., de la  
Bastide, M., Huang, E.N., King, L., Kirchoff, K.A., Miller, B.,  
Nascimento, L.U., O'Shaughnessy, A.L., Preston, R.R., Rodriguez, M.A.,  
Shah, R.S., Shekher, M., Spiegel, L.A., Toth, K. and Vill, M.D.  
TITLE Mouse Genomic Sequence  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 193432)  
TITLE Direct Submission  
JOURNAL McCombie, W.R.  
AUTHORS Submitted (13-JUN-2000) Lita Annenberg Hazen Genome Sequencing  
JOURNAL Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring  
Harbor, NY 11724, USA  
COMMENT ----- Genome Center  
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor  
Laboratory  
Laboratory Center code: CSHL  
Center code: [www.cshl.org/genseq](http://www.cshl.org/genseq)  
Web site: [www.cshl.org/genseq](http://www.cshl.org/genseq)  
Contact: [mccombie@cshl.org](mailto:mccombie@cshl.org)  
----- Project Information  
Center project name: RP23-441D12  
Center clone name: RP23-441D12  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 29 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 22334: contig of 22334 bp in length  
22335 22356: gap of unknown length  
22357 44196: contig of 21840 bp in length  
44197 44218: gap of unknown length  
44219 56959: contig of 12741 bp in length  
56960 56981: gap of unknown length  
56982 69481: contig of 12500 bp in length  
69482 69503: gap of unknown length  
69504 79748: contig of 10245 bp in length  
79749 79770: gap of unknown length  
79771 89622: contig of 9852 bp in length  
89623 89644: gap of unknown length  
89645 98721: contig of 9077 bp in length  
98722 98743: gap of unknown length  
98744 107773: contig of 9030 bp in length  
107774 107795: gap of unknown length  
107796 116484: contig of 8689 bp in length  
116485 116506: gap of unknown length  
116507 124230: contig of 7724 bp in length  
124231 124252: gap of unknown length  
124253 131603: contig of 7350 bp in length  
131603 131623: gap of unknown length  
131624 138845: contig of 7222 bp in length  
138846 138866: gap of unknown length

138867 146038: contig of 7172 bp in length  
146039 146059: gap of unknown length  
146060 150938: contig of 4879 bp in length  
150939 150959: gap of unknown length  
150960 155607: contig of 4648 bp in length  
155608 155628: gap of unknown length  
155629 159754: contig of 4126 bp in length  
159755 159775: gap of unknown length  
159776 163798: contig of 4023 bp in length  
163799 163819: gap of unknown length  
163820 167684: gap of unknown length  
167685 171445: contig of 3740 bp in length  
171446 171465: gap of unknown length  
171466 174891: contig of 3426 bp in length  
174892 174912: gap of unknown length  
174913 177626: contig of 2714 bp in length  
177627 177647: gap of unknown length  
177648 180331: contig of 2684 bp in length  
180332 180352: gap of unknown length  
180353 182820: contig of 2468 bp in length  
182821 182841: gap of unknown length  
182842 184811: contig of 1970 bp in length  
184812 184832: gap of unknown length  
184833 186797: contig of 1965 bp in length  
186798 186818: gap of unknown length  
186819 188702: contig of 1884 bp in length  
188703 188723: gap of unknown length  
188724 190562: contig of 1839 bp in length  
190563 190583: gap of unknown length  
190584 192354: contig of 1771 bp in length  
192355 192375: gap of unknown length  
192376 193432: contig of 1057 bp in length.

FEATURES  
Location/Qualifiers  
1..193432  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-441D12"

BASE COUNT 53686 a 44047 c 43159 g 51814 t 726 others  
ORIGIN

Query Match 10.1%; Score 96.8; DB 57; Length 193432;  
Best Local Similarity 74.5%; Pred. No. 2e-14;  
Matches 149; Conservative 0; Mismatches 47; Indels 4; Gaps 2;

Qy 266 gtaccgagtggtgggggtgctacagccagcgttctcaacctctctgatgctcg 325  
Db 53339 GTACCCACAAATTGAAGAAGCTGGGTCTAGAGTTGTGTTCTCAACCTTCTCTATGTCGCG 53398

Qy 326 accctttaacacagtcctctgctggtgacctcccccaaccttaaaattattttgtt 385  
Db 53399 ACCCTTTAATACAA--CTCCTTTGTTGTTGTGACCTCAACCAATAAATATTTTGT 53455

Qy 386 gct-gttcataactgtgattttgatctactgttatgaataataataattttgaaga 444  
Db 53456 GCTACTTCAATCACTGTAATTTTGTCTACTGTATGAATGTAATGTGAGTATCTTTTGAGA 53515

Qy 445 aagaggtttgccaggggttt 464  
Db 53516 TAGAGGTTTGTTCAGGGGGTT 53535

RESULT 7  
LOCUS AR016448 2830 bp DNA PAT 05-DEC-1998  
DEFINITION Sequence 3 from patent US 5776731.  
ACCESSION AR016448  
VERSION AR016448.1 GI:3972725  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.





Db 187101 TCTGAATACGTCAGGGGCTACAA-----GTAGCCATCGGACTATCTCC 187145  
QY 820 tggatgaaggaggaagtcagttctctcttggggaagtttgcattctgtgtaacgatatct 879  
Db 187146 TGTGATT-AGGAGTCTGGGTTCTTGGGGCAAGTTTGATCTAGCTGTCACAAATCT 187204  
QY 880 aattctctccctgttctgtctgtctgttgaacaacgactgtgatacccaacaatggacc 939  
Db 187205 ACTTTTTC----TTTGTCTTCTCTTGTGCGAGGACTACTTAACAACCACTGCGCAACA 187261  
QY 940 atcaacca 947  
Db 187262 ACAACCAA 187269

RESULT 10  
AC016087  
LOCUS  
DEFINITION Homo sapiens clone RP11-24K13, WORKING DRAFT SEQUENCE, 38 unordered  
pieces.  
ACCESSION AC016087  
VERSION AC016087.5 GI:6957544  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 178016)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens, clone RP11-24K13  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 178016)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barnes,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 11, 2000 this sequence version replaced gi:6957531.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L484  
Center clone name: 24\_K13  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 154757 bases at least Q40  
Consensus quality: 168173 bases at least Q30  
Consensus quality: 173336 bases at least Q20  
Insert size: 155000; agarose-fp  
Insert size: 178016; sum-of-contigs  
Quality coverage: 3.5 in Q20 bases; agarose-fp  
Quality coverage: 3.1 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 38 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1024: contig of 1024 bp in length  
\* 1025: gap of unknown length  
\* 3090: contig of 2066 bp in length  
\* 3091: gap of unknown length  
\* 4766: contig of 1676 bp in length  
\* 4767: gap of unknown length  
\* 6125: contig of 1359 bp in length  
\* 6126: gap of unknown length  
\* 7637: contig of 1512 bp in length  
\* 7638: gap of unknown length  
\* 9626: contig of 1988 bp in length  
\* 9627: gap of unknown length  
\* 10106: contig of 481 bp in length  
\* 10107: gap of unknown length  
\* 11709: contig of 1603 bp in length  
\* 11710: gap of unknown length  
\* 14917: contig of 3208 bp in length  
\* 14918: gap of unknown length  
\* 15709: contig of 792 bp in length  
\* 15710: gap of unknown length  
\* 18239: contig of 2530 bp in length  
\* 18240: gap of unknown length  
\* 20913: contig of 2674 bp in length  
\* 20914: gap of unknown length  
\* 25237: contig of 4324 bp in length  
\* 25238: gap of unknown length  
\* 27396: contig of 2159 bp in length  
\* 27397: gap of unknown length  
\* 29822: contig of 2426 bp in length  
\* 29823: gap of unknown length  
\* 33125: contig of 3303 bp in length  
\* 33126: gap of unknown length  
\* 37547: contig of 4422 bp in length  
\* 37548: gap of unknown length  
\* 39975: contig of 2428 bp in length  
\* 39976: gap of unknown length  
\* 42996: contig of 3021 bp in length  
\* 42997: gap of unknown length  
\* 46556: contig of 3560 bp in length  
\* 46557: gap of unknown length  
\* 52573: contig of 6017 bp in length  
\* 52574: gap of unknown length  
\* 58459: contig of 5886 bp in length  
\* 63653: contig of 5194 bp in length  
\* 63654: gap of unknown length  
\* 69130: contig of 5477 bp in length  
\* 69131: gap of unknown length  
\* 75009: contig of 5879 bp in length  
\* 75010: gap of unknown length  
\* 80148: contig of 5139 bp in length  
\* 80149: gap of unknown length  
\* 86309: contig of 6161 bp in length  
\* 86310: gap of unknown length  
\* 93469: contig of 7160 bp in length  
\* 93470: gap of unknown length  
\* 99882: contig of 5513 bp in length  
\* 99883: gap of unknown length  
\* 104921: contig of 5939 bp in length  
\* 104922: gap of unknown length  
\* 112526: contig of 7605 bp in length  
\* 112527: gap of unknown length  
\* 122037: contig of 9511 bp in length  
\* gap of unknown length



122038 130295: contig of 8258 bp in length  
gap of unknown length  
130296 137837: contig of 7542 bp in length  
gap of unknown length  
137838 148168: contig of 10331 bp in length  
gap of unknown length  
148169 154467: contig of 6299 bp in length  
gap of unknown length  
154468 164476: contig of 10009 bp in length  
gap of unknown length  
164477 178016: contig of 13540 bp in length.

FEATURES  
Source  
1. 178016  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-24K13"  
/clone\_lib="RPCI-11 Human Male BAC"  
BASE COUNT 51839 a 36865 c 36521 g 52752 t 39 others  
ORIGIN

Query Match 9.9%; Score 95.2; DB 39; Length 178016;  
Best Local Similarity 77.5%; Pred. NO. 5.2e-14;  
Matches 141; Conservative 0; Mismatches 38; Indels 3; Gaps 2;

QY 286 tgcctacagccaggggtctcaacctctcgtatgcttcgaccccttaataca--gtgcc 343  
DB 147557 TGCTTTAAACAGGGTCTCAACCTCTTATGTTGGCTTTTATACATGTTCT 147616

QY 344 tcatgtctgtgaccccccacacattttttgttgc-tgtcataactgtga 402

DB 147617 TCATGTTATGGTGACCTTCACACAAATATTTTGTCTATTATATATAA 147676

QY 403 tttgtactgttatgaattgtaataataattttgaagaagaggttgcgaaggt 462

DB 147677 TTTGCTACTGTTATGAATATAGTATAAATATTTTGAGATAGAGTTGCCAAAAGT 147736

QY 463 tt 464

DB 147737 GT 147738

RESULT 11  
AC073759/c  
DEFINITION Mus musculus clone RP23-33G12, WORKING DRAFT SEQUENCE, 59 unordered  
pieces.  
AC073759  
VERSION AC073759.1 GI:8810376  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE 1 (bases 1 to 314146)

JOURNAL Sequencing of Mouse  
AUTHORS DOE Joint Genome Institute.  
TITLE 2 (bases 1 to 314146)

JOURNAL Direct Submission  
AUTHORS DOE Joint Genome Institute.  
TITLE Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 1749700  
Center clone name: RPCI-23\_33G12  
-----  
Summary Statistics

Consensus quality: 272713 bases at least Q40  
Consensus quality: 292613 bases at least Q30  
Consensus quality: 296870 bases at least Q20  
Estimated insert size: 245000; agarose-fp estimation  
Quality coverage: 7.75 in Q20 bases; agarose-fp estimation  
Quality coverage: 6.16 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 59 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1127: contig of 1127 bp in length  
1227: gap of unknown length  
1228 2405: contig of 1178 bp in length  
1228 2505: gap of unknown length  
2406 4058: contig of 1553 bp in length  
2506 4158: gap of unknown length  
4059 5224: contig of 1066 bp in length  
4159 5324: gap of unknown length  
5325 6598: contig of 1174 bp in length  
5325 6598: gap of unknown length  
6499 7847: contig of 1249 bp in length  
6599 7948: gap of unknown length  
7848 9144: contig of 1197 bp in length  
7948 9145: gap of unknown length  
9145 10879: contig of 1635 bp in length  
9245 10880: gap of unknown length  
10880 12128: contig of 1149 bp in length  
10980 12228: gap of unknown length  
12129 13473: gap of unknown length  
12229 13572: gap of unknown length  
13473 14606: contig of 1034 bp in length  
13573 14706: gap of unknown length  
14607 15750: contig of 1044 bp in length  
14707 15751: gap of unknown length  
15751 17125: contig of 1275 bp in length  
15851 17225: gap of unknown length  
17126 18483: contig of 1258 bp in length  
17226 18584: gap of unknown length  
18484 20388: contig of 1805 bp in length  
20389 20488: gap of unknown length  
20489 21780: gap of unknown length  
21781 23376: contig of 1596 bp in length  
21781 23377: gap of unknown length  
23377 24606: contig of 1130 bp in length  
24607 24706: gap of unknown length  
24707 25710: contig of 1004 bp in length  
25711 26865: contig of 1055 bp in length  
25811 26966: gap of unknown length  
26866 28143: contig of 1178 bp in length  
28144 28243: gap of unknown length  
28244 29698: contig of 1455 bp in length  
29699 31603: gap of unknown length  
31603 31703: contig of 1805 bp in length  
31704 33388: contig of 1685 bp in length  
33389 34828: gap of unknown length  
34829 36903: contig of 1340 bp in length  
36904 37003: gap of unknown length  
37004 39194: gap of unknown length  
39195 41491: contig of 2296 bp in length  
41491 42966: gap of unknown length  
42967 43066: contig of 1376 bp in length  
43067 43066: gap of unknown length

```

* 43067 45019: contig of 1953 bp in length
* 45020 45119: gap of unknown length
* 45120 45164: contig of 1345 bp in length
* 46465 46564: gap of unknown length
* 46565 48212: contig of 1548 bp in length
* 48213 48312: gap of unknown length
* 48313 48445: contig of 1533 bp in length
* 48446 49945: gap of unknown length
* 49946 52230: contig of 2285 bp in length
* 52231 52330: gap of unknown length
* 52331 54159: contig of 1829 bp in length
* 54160 54259: gap of unknown length
* 54260 55781: contig of 1522 bp in length
* 55782 55881: gap of unknown length
* 55882 58225: contig of 2344 bp in length
* 58226 58325: gap of unknown length
* 58326 60141: contig of 1816 bp in length
* 60142 60241: gap of unknown length
* 60242 62745: contig of 2504 bp in length
* 62746 62845: gap of unknown length
* 62846 66534: contig of 3689 bp in length
* 66535 66634: gap of unknown length
* 66635 70286: contig of 3652 bp in length
* 70287 70386: gap of unknown length
* 70387 77088: contig of 6701 bp in length
* 77088 77187: gap of unknown length
* 77188 81368: contig of 4181 bp in length
* 81369 81468: gap of unknown length
* 81469 86840: contig of 5372 bp in length
* 86841 86940: gap of unknown length
* 86941 96600: contig of 9660 bp in length
* 96601 96700: gap of unknown length
* 96701 106003: contig of 9303 bp in length
* 106004 106103: gap of unknown length
* 106104 114533: contig of 8429 bp in length
* 114533 127575: contig of 12943 bp in length
* 127576 127675: gap of unknown length
* 127676 141198: contig of 13523 bp in length
* 141199 141298: gap of unknown length
* 141299 152661: contig of 11363 bp in length
* 152662 152761: gap of unknown length
* 152762 163299: contig of 10538 bp in length
* 163300 163399: gap of unknown length
* 163400 174025: contig of 10626 bp in length
* 174125 174125: gap of unknown length
* 174126 187824: contig of 13699 bp in length
* 187825 187924: gap of unknown length
* 187925 200427: contig of 12503 bp in length
* 200428 200527: gap of unknown length
* 200528 217401: contig of 16874 bp in length
* 217402 217501: gap of unknown length
* 217502 232676: contig of 15175 bp in length
* 232677 232776: gap of unknown length
* 232777 253694: contig of 20918 bp in length
* 253695 253794: gap of unknown length
* 253795 279819: contig of 26025 bp in length
* 279820 279919: gap of unknown length
* 279920 314146: contig of 34227 bp in length.
* Location/Qualifiers
  1. .314146
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone_lib="RP23-33G12"
    /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 77948 a 78262 c 78357 g 73694 t 5885 others
ORIGIN

```

## FEATURES

```

source
1. .314146
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone_lib="RP23-33G12"
  /clone_lib="RPCI mouse BAC library 23"

```

```

BASE COUNT 77948 a 78262 c 78357 g 73694 t 5885 others
ORIGIN

```

```

Query Match 9.9% Score 94.6; DB 57; Length 314146;
Best Local Similarity 66.1%; Pred. No. 7.7e-14;
Matches 191; Conservative 0; Mismatches 79; Indels 19; Gaps 3;

```

```

QY 640 tggagtttggtacttggtggtctcttttttttcaaaccttttttactcttttttccacc 599
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55452 TGTCTTTTGTCTTTTGTGGTCTTTCTTTTCATCATCTTCACCCCTTTTCTTCAACC 55393
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 700 tgtcgcccccctaacactaaataagaagaagaaggggagcatagaggggaaagaaacc 759
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55392 TTTTAAACCCCTTAACACTAGATAAGAGAGAAGAGGATAGAGGAAAGGATGAGATC 55333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 760 cctgaataacytcagtagttggcaaggggggtgacatatgtgtcatttagaccacatcc 819
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55332 TCTGAATAACGTCAGGGGCTACAAA-----GTAGCCATCGGACTACTTCC 55288
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 820 tggtagtaaygggagtcgaattcccttgaggcaagtttgatcttctgtaacgatatct 879
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55287 TCGTGATT-AGGAGTGTGCGGTTCTTTGGGCAAGTTTGATCTTAGCTGTCFAAAATCT 55229
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 880 aatttctctctgtgtctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 928
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55228 ACTTTTCTCT-TTTTGTTCTTCTTTGTGCGAGGACTACTTAAACAACC 55183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

## RESULT 12

## AC068140

## LOCUS

## DEFINITION

## AC068140

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: [nisc\\_mouse@hgrl.nih.gov](mailto:nisc_mouse@hgrl.nih.gov)

----- Project Information

Center project name: xm

Center clone name: 246M15

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 184080 bases at least Q40

Consensus quality: 186055 bases at least Q30

Consensus quality: 187380 bases at least Q20

Insert size: 197000; agarose-fp

Insert size: 184000; pulse-field-gel

Quality coverage: 6.57x in Q20 bases; agarose-fp

Quality coverage: 7.03x in Q20 bases; pulse-field-gel

Quality coverage: 6.76x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 11 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 3556: contig of 3556 bp in length  
 gap of unknown length  
 3557 7625: contig of 4069 bp in length  
 gap of unknown length  
 7626 12387: contig of 4762 bp in length  
 gap of unknown length  
 12388 17199: contig of 4812 bp in length  
 gap of unknown length  
 17200 22561: contig of 5362 bp in length  
 gap of unknown length  
 22562 33699: contig of 11138 bp in length  
 gap of unknown length  
 33700 48936: contig of 15237 bp in length  
 gap of unknown length  
 48937 77595: contig of 28659 bp in length  
 gap of unknown length  
 77596 112463: contig of 34868 bp in length  
 gap of unknown length  
 112464 141437: contig of 28974 bp in length  
 gap of unknown length  
 141438 191276: contig of 49839 bp in length.

## FEATURES

Location/Qualifiers  
 1..191276  
 /organism="Mus musculus"  
 /strain="C57BL6/J"  
 /db\_xref="taxon:10090"  
 /chromosome="5"  
 /clone="RP23-246M15"  
 /clone\_lib="RPCI mouse BAC library 23"  
 BASE COUNT 56482 a 39792 c 38890 g 56096 t 16 others  
 ORIGIN

Query Match 9.8%; Score 94; DB 55; Length 191276;

Best Local Similarity 78.4%; Pred. NO. 1.le-13;

Matches 138; Conservative 0; Mismatches 33; Indels 3; Gaps 2;

QY 288 cctacagcagggttcacaccttcctgctgcaccctttaaatacagctgcctcat 347

Db 108532 CATCTAGACGATGGTCTCAACCTCTCAATGATGCTACTTTTGATATCTCCCTCAT 108591

QY 348 gctctggtgacctcccaaccttaaaattattttgtgct-gttcataactgtgatttt 406

Db 108592 ATTGTGCTGAC--CCACAACATATAAATATTTTCTGCTACTTCTGCTGTAAATTT 108649

QY 407 gatactgttatgaattgataataaattttgaagaagaggtttgccagggt 462

Db 108650 GCTACTGTTATGAATGTAATGTAATATTTTGGAGCTAGAGGTTTGTAAAGGT 108705

## RESULT 13

AC044778 195190 bp DNA HTG 21-JUL-2000  
 LOCUS Mus musculus clone RP23-34M17 strain C57BL6/J, WORKING DRAFT  
 DEFINITION SEQUENCE, 36 unordered pieces.

AC044778

AC044778.7 GI:9309499

ACCESSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.

VERSION house mouse.

KEYWORDS Mus musculus

SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 195190)

AUTHORS Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,

Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,

Gordon, M., Goltz, J.S. and Kucherlapati, R.

TITLE High Throughput Mouse Sequencing

JOURNAL Unpublished

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

2 (bases 1 to 195190)  
 Han, J., Montgomery, K.T., Grills, G., Lee, E., Long, J., Pomerantz, R.,  
 Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,  
 Gordon, M., Goltz, J.S. and Kucherlapati, R.  
 Direct submission  
 Submitted (12-APR-2000) Department of Molecular Genetics, Albert  
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
 Bronx, NY 10461, USA  
 On Jul 21, 2000 this sequence version replaced gi:7798714.  
 -----Genome Center Center:  
 Albert Einstein College of Medicine Center:  
 Code: #ECOM  
 Web site:  
 http://sequence.aecom.yu.edu/cgi-  
 bin/ws.exe/mouseDB/mouseSEQ/mouseseqtable.hts  
 Contact: jhan@sequence.aecom.yu.edu  
 -----Summary Statistics

Center project name: ACB  
 Sequencing vector: pUC18; L08752  
 Chemistry: Dye-terminator Big Dye; 100%  
 \*Consensus quality: 180994 at least Q20  
 \*Consensus quality: 174798 at least Q30  
 \*Consensus quality: 165101 at least Q40  
 \*Estimated insert size: agarose-PP - N/A  
 \*Estimated insert size: 194490 - sum-of-contigs  
 Quality coverage: agarose-PP - N/A  
 Quality coverage: 4.7x sum-of-contigs - N/A

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 36 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 11742: contig of 11742 bp in length  
 \* 11743 11762: gap of unknown length  
 \* 11763 12029: contig of 267 bp in length  
 \* 12030 12049: gap of unknown length  
 \* 12050 12297: contig of 10858 bp in length  
 \* 22308 22927: gap of unknown length  
 \* 33305 33325: contig of 10378 bp in length  
 \* 33326 41564: contig of 8239 bp in length  
 \* 41565 50140: contig of unknown length  
 \* 50141 50160: gap of unknown length  
 \* 50161 58861: contig of 8701 bp in length  
 \* 58862 66717: contig of 7836 bp in length  
 \* 66718 66737: gap of unknown length  
 \* 66738 74346: contig of 7609 bp in length  
 \* 74347 74366: gap of unknown length  
 \* 74367 82536: contig of 8170 bp in length  
 \* 82537 92171: contig of 9615 bp in length  
 \* 92172 92191: gap of unknown length  
 \* 92192 99077: contig of 6886 bp in length  
 \* 99078 99097: gap of unknown length  
 \* 99098 106290: contig of 7192 bp in length  
 \* 106291 106309: gap of unknown length  
 \* 106310 112038: contig of 5729 bp in length  
 \* 112039 112058: gap of unknown length  
 \* 112059 117815: contig of 5757 bp in length  
 \* 117816 117835: gap of unknown length  
 \* 117836 122969: contig of 5134 bp in length  
 \* 122970 122989: gap of unknown length  
 \* 122990 129847: contig of 6858 bp in length  
 \* 129848 129867: gap of unknown length  
 \* 129868 134558: contig of 4691 bp in length  
 \* 134559 134578: gap of unknown length  
 \* 134579 140199: contig of 5621 bp in length

```
* 140200 140219: gap of unknown length
* 140220 145726: contig of 5507 bp in length
* 145727 145746: gap of unknown length
* 145747 149364: contig of 3618 bp in length
* 149365 149384: gap of unknown length
* 149385 153542: contig of 4158 bp in length
* 153543 153562: gap of unknown length
* 153563 156634: contig of 3072 bp in length
* 156635 156654: gap of unknown length
* 156655 160471: contig of 3817 bp in length
* 160472 160491: gap of unknown length
* 160492 165611: contig of 5120 bp in length
* 165612 165631: gap of unknown length
* 165632 168936: contig of 3305 bp in length
* 168937 172409: gap of unknown length
* 172410 172429: gap of unknown length
* 172430 175364: contig of 2935 bp in length
* 175365 175384: gap of unknown length
* 175385 178815: contig of 3431 bp in length
* 178816 178835: gap of unknown length
* 178836 181876: contig of 3041 bp in length
* 181877 181896: gap of unknown length
* 181897 184423: contig of 2527 bp in length
* 184424 184443: gap of unknown length
* 184444 186464: contig of 2021 bp in length
* 186465 186484: gap of unknown length
* 186485 186808: contig of 1124 bp in length
* 186809 188628: gap of unknown length
* 188629 190822: contig of 2194 bp in length
* 190823 190842: gap of unknown length
* 190843 192888: contig of 2046 bp in length
* 192889 192908: gap of unknown length
* 192909 195190: contig of 2282 bp in length.

FEATURES             Location/Qualifiers
    source            1..195190
                    /organism="Mus musculus"
                    /strain="C57BL6/J"
                    /db_xref="taxon:10090"
                    /clone="RP23-34M17"
                    /sex="male"

BASE COUNT   51708 a 45497 g 45097 t 918 others
ORIGIN

Query Match
Best Local Similarity 9.7% Score 93.4; DB 53; Length 195190;
Matches 150; Conservative 0; Mismatches 36; Indels 5; Gaps 3;

QY 287 gcctacagccagcgggtctcaacctctctgctgcttgcaccccttttaatacagtgctca 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133004 CCCCACTGAGCAGTGGTCTCAACCTCTGCTGCTATGACCTTTAATACAGTTCTCA 133063

QY 347 tgcctggtgacctcccaaccttaatttttttctt-gttcataactgtgattt 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133064 TGCTGTGGTAA---CCCAACCAATAATTTATTTTGTCTACTTCAATACCTGTAACT 133120

QY 406 tgatactgttatgaattgtaataataataatttgaagaagaggttgcacagggt-tt 464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133121 TGCTAATATTATGATCATATGTAATATTTTGGAGACAGAGGTTCCAAAGAGTCAT 133180

QY 465 gagaactgctg 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133181 GAGAACTGCTG 133191

RESULT 14
LOCUS      AC016464      207160 bp      DNA      HTG      01-MAR-2000
DEFINITION Mus musculus chromosome 11 clone RP23-409J21 map 11, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION AC016464
VERSION   AC016464.3 GI:7137319
```

KEYWORDS  
SOURCE  
ORGANISMREFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORSTITLE  
JOURNAL  
COMMENT

HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 207160)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Mus musculus chromosome 11, clone RP23-409J21  
Unpublished  
2 (bases 1 to 207160)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melidrim,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (30-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 1, 2000 this sequence version replaced gi:6970326.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://fpc.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L3198  
Center clone name: 409\_JL21  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 196848 bases at least Q40  
Consensus quality: 202204 bases at least Q30  
Consensus quality: 204082 bases at least Q20  
Insert size: 210000; agarose-fp  
Quality coverage: 4.8 in Q20 bases; agarose-fp  
Quality coverage: 4.9 in Q20 bases; sum-of-contigs

\* NOTE: this is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* As soon as it is available and the accession number will  
\* be preserved.

1 541: contig of 541 bp in length  
\* 542 641: gap of 100 bp  
\* 642 1859: contig of 1218 bp in length  
\* 1860 1959: gap of 100 bp  
\* 1960 4152: contig of 2193 bp in length  
\* 4153 4252: gap of 100 bp  
\* 4253 6291: contig of 2039 bp in length  
\* 6292 6391: gap of 100 bp  
\* 6392 10537: contig of 4146 bp in length  
\* 10538 10637: gap of 100 bp  
\* 10638 14144: contig of 3507 bp in length  
\* 14145 14244: gap of 100 bp  
\* 14245 18473: contig of 4235 bp in length  
\* 18480 18579: gap of 100 bp

```

* 18580 25538: contig of 6959 bp in length
* 25539 25638: gap of 100 bp
* 25639 31995: contig of 6357 bp in length
* 31996 32095: gap of 100 bp
* 32096 40133: contig of 8038 bp in length
* 40134 40233: gap of 100 bp
* 40234 48264: contig of 8031 bp in length
* 48265 48364: gap of 100 bp
* 48365 55351: contig of 6987 bp in length
* 55352 55451: gap of 100 bp
* 55452 63846: contig of 8395 bp in length
* 63847 63946: gap of 100 bp
* 63947 75458: contig of 11510 bp in length
* 75459 75557: gap of 100 bp
* 75558 88148: contig of 12592 bp in length
* 88149 88248: gap of 100 bp
* 88249 99481: contig of 11233 bp in length
* 99482 99581: gap of 100 bp
* 99582 115973: contig of 16392 bp in length
* 115974 116073: gap of 100 bp
* 116074 132294: contig of 16221 bp in length
* 132295 132394: gap of 100 bp
* 132395 149262: contig of 16868 bp in length
* 149263 149362: gap of 100 bp
* 149363 177577: contig of 28215 bp in length
* 177578 177677: gap of 100 bp
* 177678 207160: contig of 29483 bp in length.

```

## FEATURES

source

```

Location/Qualifiers
1..207160
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/map="11"
/clone="RP23-409J21"
/db_lib="RPC1-23 Female Mouse BAC"
1..541
/note="assembly_fragment"
clone_end:77
vector_side:right
642..1859
/note="assembly_fragment"
1960..4152
/note="assembly_fragment"
4253..6291
/note="assembly_fragment"
6392..10537
/note="assembly_fragment"
10638..14144
/note="assembly_fragment"
14245..18479
/note="assembly_fragment"
18580..25538
/note="assembly_fragment"
25639..31995
/note="assembly_fragment"
32096..40133
/note="assembly_fragment"
40234..48264
/note="assembly_fragment"
48365..55351
/note="assembly_fragment"
55452..63846
/note="assembly_fragment"
63947..75456
/note="assembly_fragment"
75557..88148
/note="assembly_fragment"
88249..99481
/note="assembly_fragment"
99582..115973
/note="assembly_fragment"
116074..132294
/note="assembly_fragment"

```

```

misc_feature
132395..149262
/note="assembly_fragment"
clone_end:SP6
vector_side:right
149363..177577
/note="assembly_fragment"
177678..207160
/note="assembly_fragment"

BASE COUNT 57689 a 46293 c 45874 g 55300 t 2004 others
ORIGIN

Query Match 9.7%; Score 93.2; DB 39; Length 207160;
Best Local Similarity 72.8%; Pred. No. 1.7e-13;
Matches 147; Conservative 0; Mismatches 53; Indels 2; Gaps 2;

QY 297 cagcggttctcaacctctctgatgcttcgaccccttaataacagtcctcctctgtgtg 356
|||||
Db 151965 CAGTGGTTCACCTCCCTTAATGCTGAGACCCCTTAATATAGTTCCTCATGCTGGTG 151906

QY 357 acctcccaaccttaaaattattttgtct-gtccataactgtgattttgatactgtt 415
|||||
Db 151905 ATCCCCCACCACCAAGATATATTTTCATGCTACATCAATGTAATTTGGCTACTG-T 151847

QY 416 atgaattgtaataataaattttgagaagaggtttgccaaaggggtttgagaactgtg 475
|||||
Db 151846 GTCAATCATATAATGTAATATCTGAGCCAGAGGCTCTGGACCCACAGGTTGAGAACCA 151787

QY 476 ttctagccccacgtggatggtt 497
|||||
Db 151786 CTTTAGTCTTACCAGCAGGGTT 151765

```

## RESULT 15

AC074042/c

LOCUS

DEFINITION

AC074042

AC074042.1

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

TITLE

JOURNAL

COMMENT

```

AC074042, 214616 bp DNA HTG 13-JUL-2000
Mus musculus chromosome 6 clone RP23-99G18 strain C57BL6/J, WORKING
DRAFT SEQUENCE, 22 unordered pieces.
AC074042
AC074042.1 GI:9087214
HTG; HTGS_PHASE1; HTGS_DRAFT.
house mcuse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214616)
Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bouffard, G.G., Dietrich, N.L., Eagle, W.O., Gupta, J., Ho, S.-L.,
Huang, M.C., Idol, J., Lee-Lin, S.-Q., Maduro, Q.L., Maduro, V.B.,
Mastrian, S.D., McCloskey, J.C., Morse, E., Ojodu, M.A., Pearson, R.,
Stantrifop, S., Summers, T.J., Thomas, J.W., Thomas, P.J.,
Tongson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A.,
Wetherby, K.D. and Green, E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 214616)
Green, E.D.
Direct Submission
Submitted (13-JUL-2000) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@hgri.nih.gov
----- Project Information
Center project name: t_j
Center clone name: 099G18
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 202414 bases at least Q40

```



us-09-101-423a-5.rge

Mon Nov 6 10:14:36 2000





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:29:46 ; Search time 254.92 Seconds  
(without alignments)  
1414.703 Million cell updates/sec

Title: US-09-101-423A-5  
Perfect score: 960  
Sequence: 1 gaggggtgtgtgcacagt.....tcaaccacccaaccat 960

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues  
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	955.2	99.5	960	18	T72789
2	96	10.0	2830	18	T88775
3	96	10.0	2830	20	X87653
4	81.4	8.5	49999	20	Z23891
5	81.4	8.5	49999	20	Z23896
6	81	8.4	29392	19	V15422
7	79.4	8.3	874	20	X57448
8	79.4	8.3	874	20	X57470
9	70.6	7.4	37940	20	Z01026
10	67.6	7.0	48974	20	X55300
11	65.6	6.8	36901	20	Z23892
12	65.6	6.8	38886	20	Z23897

13	62.6	6.5	16442	18	X83006
14	62.2	6.5	1678	17	T34611
15	62	6.5	892	18	V06155
16	56	5.8	4164	19	V12216
17	55	5.7	44576	21	Z61522
18	49.6	5.2	893	20	V81229
19	49.6	5.2	2247	21	A10406
20	49.6	5.2	29604	18	X83005
21	48.6	5.1	5887	17	T65001
22	47	4.9	1083	20	X57479
23	46.6	4.9	1094	20	X57480
24	46.6	4.9	1107	20	X57480
25	46.6	4.9	1107	20	X57431
26	46.4	4.8	7162	19	V22214
27	46	4.8	1059	20	X57481
28	44	4.6	5305	20	Z28289
29	43.4	4.5	960	18	T72789
30	41.6	4.3	10614	16	Q89555
31	41	4.3	6645	17	T65002
32	39.8	4.1	26698	17	T08126
33	39.8	4.1	26700	15	O67902
34	39.8	4.1	26700	19	V15946
35	39.8	4.1	26700	20	V81283
36	37.6	3.9	4322	19	V58192
37	36.8	3.8	11234	20	X80991
38	36.4	3.8	3162	16	T02947
39	35.6	3.7	1641	21	Z50920
40	35.2	3.7	278	17	T09827
41	35.2	3.7	3395	20	X80983
42	34.8	3.6	387	20	X33670
43	34.6	3.6	18613	18	V74423
44	34.2	3.6	1230025	20	X91990
45	33.4	3.5	509	19	V17664

## ALIGNMENTS

RESULT 1  
T72789  
ID T72789 standard; DNA; 960 BP.  
AC T72789;  
DT T72789;  
XX T72789;  
XX 22-SEP-1997 (first entry)  
DE Metastasis inducing DNA C12.  
XX Metastasis-inducing DNA; Met-DNA; cancer; diagnosis;  
KW Metastasis-inducing DNA; Met-DNA; cancer; diagnosis;  
XX osteopontin; ss.  
XX Homo sapiens.  
XX WO9725443-A1.  
XX 17-JUL-1997.  
XX 10-JAN-1997; 97WO-GB00074.  
XX 10-JAN-1996; 96GB-0000470.  
XX (UYLI-) UNIV LIVERPOOL.  
XX Barracough BR., Rudland PS;  
XX WPI; 1997-372878/34.  
XX New isolated metastasis-inducing DNA - used to develop products to identify and treat patients at risk from metastatic tumours  
XX Claim 12; Page 27; 38pp; English.  
XX Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12

Partial mouse WRN  
Probe for detectin  
Viral infection ge  
Mouse retinoid met  
Cosmid CVO14 conta  
Murine pituitary-d  
Murine cardiac ank  
Partial mouse WRN  
Mouse cell cycle r  
Rat U3 gene trap d  
Rat U3 gene trap d  
Rat U3 gene trap d  
Rat U3 gene trap d  
Mouse Cdx2 genomic  
Rat U3 gene trap d  
Rat U3 gene trap d  
Metastasis inducin  
Hamster cholestero  
Mouse cell cycle r  
Mouse syndecan-1 g  
Syndecan gene. Mu  
Mouse syndecan-1 e  
Mouse Bsk receptor  
Sequence of plasm  
EPH-like receptor  
Human Protease and  
Tomato genomic DNA  
Sequence of plasm  
DNA tandem nucleot  
Staphylococcus aur  
Nucleotide sequenc  
Internal nucleotid

CC and c20 (T72785-90) are entirely novel short stretches of human  
 CC regulatory DNA capable of inducing metastasis. They were  
 CC identified using a method for detecting Met-DNA that involves  
 CC transferring human DNA from a malignant, metastatic cancer cell  
 CC (in this case, breast cancer) into a cell line (pref. rat Rama 37)  
 CC that produces only benign, non-metastasing tumours when injected  
 CC into a syngeneic animal, injecting those animals having metastasing  
 CC syngeneic animal, selecting those animals having metastasing  
 CC tumours, and recovering the Met-DNA from them. The isolated  
 CC Met-DNAs can be used to develop products to identify and treat  
 CC patients at risk from metastatic tumours.  
 XX  
 SQ Sequence 960 BP; 256 A; 193 C; 233 G; 278 T; 0 other;

Query Match 99.5%; Score 955.2; DB 18; Length 960;  
 Best Local Similarity 99.7%; Pred. No. 4.7e-273;  
 Matches 957; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gagggggtggtgacagattatgttttttaggaaggggttcattgaacctcagcagact 60  
 DB 1 gagggggtggtgacagattatgttttttaggaaggggttcattgaacctcagcagact 60  
 QY 61 cggtttcagaaatttaaacccctgaggggaatttttttaaatcgtatgaattctgac 120  
 DB 61 cggtttcagaaatttaaacccctgaggggaatttttttaaatcgtatgaattctgac 120  
 QY 121 atgagaaaaacagatcagaaacgttctgtgttcagaaaaaggaaggtgtgagctaa 180  
 DB 121 atgagaaaaacagatcagaaacgttctgtgttcagaaaaaggaaggtgtgagctaa 180  
 QY 181 cagactgcacactggtgttcagggcacatctggatcacaggaggtcagataatgtcccc 240  
 DB 181 cagactgcacactggtgttcagggcacatctggatcacaggaggtcagataatgtcccc 240  
 QY 241 aaaggttaaatgatttctgtgcacagtcacaggtgtgtgtgtgtgtgtgtgtgtgtgt 300  
 DB 241 aaaggttaaatgatttctgtgcacagtcacaggtgtgtgtgtgtgtgtgtgtgtgtgt 300  
 QY 301 ggttcaacctctgtatgt 360  
 DB 301 ggttcaacctctgtatgt 360  
 QY 361 ccccaaccttaaaattttttgt 420  
 DB 361 ccccaaccttaaaattttttgt 420  
 QY 421 ttgtataataataattttgaaagaggtttgtcccaaggtttgagaactgctgttcta 480  
 DB 421 ttgtataataataattttgaaagaggtttgtcccaaggtttgagaactgctgttcta 480  
 QY 481 gcccacgtgagttgttttgcatttgcatttgcatttgcatttgcatttgcatttgcattt 540  
 DB 481 gcccacgtgagttgttttgcatttgcatttgcatttgcatttgcatttgcatttgcattt 540  
 QY 541 caggctagcagcttagatgtgctacttagctgaggaataacaccttgaaactctgtgagac 600  
 DB 541 caggctagcagcttagatgtgctacttagctgaggaataacaccttgaaactctgtgagac 600  
 QY 601 tggagagactggttagctctcagaagactggaataagctggagtttggctacttgggg 660  
 DB 601 tggagagactggttagctctcagaagactggaataagctggagtttggctacttgggg 660  
 QY 661 ttcttttttcaaaccttttctactcttttccacctgtggccctccttaacactaa 720  
 DB 661 ttcttttttcaaaccttttctactcttttccacctgtggccctccttaacactaa 720  
 QY 721 taagaaagagaaaggggagcatagagggggaaagaaacccctgaataaagctagtggtg 780  
 DB 721 taagaaagagaaaggggagcatagagggggaaagaaacccctgaataaagctagtggtg 780  
 QY 781 gcaaggggggtgacatatgttgcattatagaccacatctcgtgtgtaaggggagtcag 840  
 DB 781 gcaaggggggtgacatatgttgcattatagaccacatctcgtgtgtaaggggagtcag 840

Db 781 gcaaggggggtgacatatgttgcattatagaccacatctcgtgtgtaaggggagtcag 840  
 QY 841 ttctttggggcaagtgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatt 900  
 Db 841 ttctttggggcaagtgttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatt 900  
 QY 901 cgttttgcagacagactgtatgaacccaatggaccatcaccacccaacccaacccaacccaac 960  
 Db 901 cgttttgcagacagactgtatgaacccaatggaccatcaccacccaacccaacccaacccaac 960

RESULT 2

T88775/C

ID T88775 standard: cDNA; 2830 BP.

XX AC T88775;

XX DT 22-FEB-1998 (first entry)

XX XX Mouse receptor protein 2F1 cDNA clone mu2F1.

DE 2F1; receptor; prostaglandin; inflammation; arthritis; allergy;

KW asthma; inflammatory bowel disease; fever; diabetes;

KW dysmenorrhoea; therapy; mouse; ss.

XX OS Mus musculus.

XX Key Location/Qualifiers

FT CDS 381..1994

FT /\*tag= a

FT sig\_peptide 381..434

FT /\*tag= b

FT mat\_peptide 435..1991

FT /\*tag= c

XX WO9731010-A1.

XX 28-AUG-1997.

XX 06-FEB-1997; 97WO-US01697.

XX 21-FEB-1996; 96US-0604333.

XX (IMMV ) IMMUNEX CORP.

XX Parnet P, Sims JE;

XX WPI: 1997-435081/40.

XX P-PSDB; W31908.

PT DNA encoding 2F1 receptors that inhibit prostaglandin synthesis -

PT for treatment of inflammatory diseases, e.g. arthritis, allergy,

PT asthma, etc.

XX Example 2; Page: 36-38; 49pp; English.

CC cDNA clone mu2F1 encodes a novel mouse receptor protein designated

CC 2F1 (see W31908) that is a member of the interleukin-1 receptor

CC (IL-1R) family. It was isolated from a murine EL4 6.1 cDNA library

CC using a probe based on human 2F1 cDNA (see 188774). The isolated

CC cDNA can be used for production of recombinant 2F1, especially

CC soluble 2F1, in transformed host cells for use in claimed methods

CC of inhibiting prostaglandin synthesis in a mammal, and of reducing

CC inflammation in a mammal (e.g. for treating arthritis, allergy,

CC asthma, inflammatory bowel disease, fever, diabetes, dysmenorrhoea,

CC etc.). Since 2F1 can inhibit NF-kappaB activity, it may also be

CC used to reduce expression of human immunodeficiency virus in

CC infected cells (not claimed). Fragments of DNA are useful as probes

CC to isolate related sequences, in hybridisation assays and as

CC sense or antisense molecules to block expression of 2F1 protein.

XX Sequence 2830 BP; 830 A; 594 C; 669 G; 737 T; 0 other;





Db 7746 ctgttatgaatcgttaataatctgatatcgagagatatctgatcgaccctgt 7805  
 Qy 466 agaaactgtcttagcccca 486  
 Db 7806 gaaagggtgttgactcca 7826

RESULT 7  
 X57448/c  
 ID X57448 standard; DNA; 874 BP.  
 XX  
 AC X57448;  
 XX  
 DT 24-JUL-1999 (first entry)  
 XX  
 DE Rat U3 gene trap derived nucleic acid 31 4 4#/rE.  
 XX  
 KW Gene trap; rat; cellular gene; viral infection; cell survival; cancer;  
 KW tumour progression; suppression; identification; viral growth;  
 KW tumour suppressor; prevention; screening; therapeutic agent; ss.  
 OS Rattus norvegicus.  
 XX  
 PN WO9919481-A2.  
 XX  
 PD 22-APR-1999.  
 XX  
 PF 08-OCT-1998; 98WO-US21276.  
 XX  
 PR 10-OCT-1997; 97US-0062021.  
 XX  
 PA (UYVA-) UNIV VANDERBILT.  
 XX  
 PI Dubois RN, Organ EL, Rubin DH;  
 XX  
 DR WPI; 1999-326546/27.  
 XX  
 PT Nucleic acid encoding tumor suppressors and products required for  
 PT viral infection  
 XX  
 PS Claim 1; Page 73-74; 94pp; English.  
 XX  
 CC This invention describes novel rat-derived nucleic acid fragments from  
 CC cellular genes that are necessary for viral infection but not for cell  
 CC survival, or that suppress tumour progression. The products of the  
 CC invention (X57371-X57497) can be used in methods of identifying cellular  
 CC genes necessary for viral growth and cellular genes that function as  
 CC tumour suppressors and for reducing or preventing such infections or  
 CC cancer. They may also be used in screening for potential therapeutic  
 CC agents. These sequences can be targeted without significant side effects  
 CC (contrast targeting genes essential for viral growth).  
 XX  
 SQ Sequence 874 BP; 207 A; 180 C; 203 G; 254 T; 30 other;

Query Match 8.3%; Score 79.4; DB 20; Length 874;  
 Best Local Similarity 73.1%; Pred. No. 7.9e-14;  
 Matches 98; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
 Qy 301 ggttctcaacctctcgtatgcttcgacctttaatacagtcgctcgtcgtgaccc 360  
 Db 393 GGTCTCAACCTTCTTAATGCTAAANCCCTGTGATACCGTCCCTCTGTGTGGTGACCC 334  
 Qy 361 ccccaaccttaaaattttttgtgtgttcataactgtgattgtgatactgttata 420  
 Db 333 TCCCAACATAGAGTATTTCATTCGTACTCCATACTGTATTTTGTCTTANGCTTAA 274  
 Qy 421 ttgtaataataa 434  
 Db 273 TCCTAATGTTAATA 260

RESULT 8  
 X57470/c  
 ID X57470 standard; DNA; 874 BP.  
 XX  
 AC X57470;  
 XX  
 DT 24-JUL-1999 (first entry)  
 XX  
 DE Rat U3 gene trap derived nucleic acid 31.4.4#-rE.  
 XX  
 KW Gene trap; rat; cellular gene; viral infection; cell survival; cancer;  
 KW tumour progression; suppression; identification; viral growth;  
 KW tumour suppressor; prevention; screening; therapeutic agent; ss.  
 OS Rattus norvegicus.  
 XX  
 PN WO9919481-A2.  
 XX  
 PD 22-APR-1999.  
 XX  
 PF 08-OCT-1998; 98WO-US21276.  
 XX  
 PR 10-OCT-1997; 97US-0062021.  
 XX  
 PA (UYVA-) UNIV VANDERBILT.  
 XX  
 PI Dubois RN, Organ EL, Rubin DH;  
 XX  
 DR WPI; 1999-326546/27.  
 XX  
 PT Nucleic acid encoding tumor suppressors and products required for  
 PT viral infection  
 XX  
 PS Claim 1; Page 82; 94pp; English.  
 XX  
 CC This invention describes novel rat-derived nucleic acid fragments from  
 CC cellular genes that are necessary for viral infection but not for cell  
 CC survival, or that suppress tumour progression. The products of the  
 CC invention (X57371-X57497) can be used in methods of identifying cellular  
 CC genes necessary for viral growth and cellular genes that function as  
 CC tumour suppressors and for reducing or preventing such infections or  
 CC cancer. They may also be used in screening for potential therapeutic  
 CC agents. These sequences can be targeted without significant side effects  
 CC (contrast targeting genes essential for viral growth).  
 XX  
 SQ Sequence 874 BP; 207 A; 180 C; 203 G; 254 T; 30 other;

Query Match 8.3%; Score 79.4; DB 20; Length 874;  
 Best Local Similarity 73.1%; Pred. No. 7.9e-14;  
 Matches 98; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 301 ggttctcaacctctcgtatgcttcgacctttaatacagtcgctcgtcgtgaccc 360  
 Db 393 GGTCTCAACCTTCTTAATGCTAAANCCCTGTGATACCGTCCCTCTGTGTGGTGACCC 334  
 Qy 361 ccccaaccttaaaattttttgtgtgttcataactgtgattgtgatactgttata 420  
 Db 333 TCCCAACATAGAGTATTTCATTCGTACTCCATACTGTATTTTGTCTTANGCTTAA 274  
 Qy 421 ttgtaataataa 434  
 Db 273 TCCTAATGTTAATA 260

RESULT 9  
 Z01026/c  
 ID Z01026 standard; DNA; 37940 BP.  
 XX  
 AC Z01026;  
 XX  
 DT 27-SEP-1999 (first entry)  
 XX

DE Partial mouse PGI genomic sequence.  
XX  
KW PGI gene; biallelic marker; PCR primer; PGI-related biallelic marker;  
KW cancer; prostate cancer; diagnosis; therapy; prostate specific antigen;  
KW PSA; mouse; ss.  
OS  
XX Mus musculus.  
XX  
PN WO9932644-A2.  
XX  
PD 01-JUL-1999.  
XX  
PF 22-DEC-1998; 98WO-IB02133.  
XX  
PR 09-SEP-1998; 98US-0099658.  
XX  
PR 22-DEC-1997; 97US-0996306.  
XX  
PA (GEST ) GENSET.  
XX  
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen D;  
XX WPI; 1999-405178/34.  
DR  
PT Use of a prostate cancer associated gene and biallelic markers  
PT derived from it  
XX  
PS Claim 3; Page 285-296; 385pp; English.  
XX  
CC The invention relates to a mammalian PGI gene and protein, and a set of  
CC PGI biallelic markers. The PGI polynucleotide and biallelic markers are  
CC used in a hybridisation assay, a sequencing assay, or in an  
CC allele-specific amplification assay for determining the identity of a  
CC nucleotide at a PGI-related biallelic marker. The methods can be used to  
CC detect and to assess the risk of developing cancer or prostate cancer.  
CC Early-stage diagnosis of prostate cancer relies on prostate specific  
CC antigen (PSA) dosage. However, the effectiveness of this is limited due  
CC to its inability to discriminate between malignant and non-malignant  
CC affections of the organ. A need exists for both a reliable diagnostic  
CC procedure which would enable early-stage diagnosis, and for preventative  
CC and curative treatments of the disease. The PGI gene can be used for  
CC detection of prostate cancer, and the risk of developing it in the  
CC future, and can also be used to determine therapies for the disease.  
XX  
SQ Sequence 37940 BP; 10347 A; 7683 C; 7920 G; 11945 T; 45 other;  
  
Query Match 7.4%; Score 70.6; DB 20; Length 37940;  
Best Local Similarity 72.1%; Pred. No. 2e-10;  
Matches 106; Conservative 0; Mismatches 39; Indels 2; Gaps 1;  
  
Qy 302 gttctcaacctctctgatgcttcgacccttaatacagtgccctcgtctgtgaccc 361  
Db 9719 GCTCTCAACCTGCTTAATCTGACCTTTAATGTCAGTCTCTACGTTATGTGAC -C 9662  
  
Qy 362 cccacaccttaaaattttttgttgccttcataactgattttgatactgttgaat 421  
Db 9661 CTCACCCATAAATATTCTTGAATGCTGTAATGCTGTTGCTGCTGTACAA 9602  
  
Qy 422 tgaataataaatttttgaagaaga 448  
Db 9601 CATAATGTAACCTATCTGTAAGCAGGA 9575  
  
RESULT 10  
X55300  
ID X55300 standard; DNA; 48974 BP.  
XX  
AC X55300;  
XX  
DT 19-JUL-1999 (first entry)  
XX  
DE Mouse Presenilin-1 gene sequence.  
XX

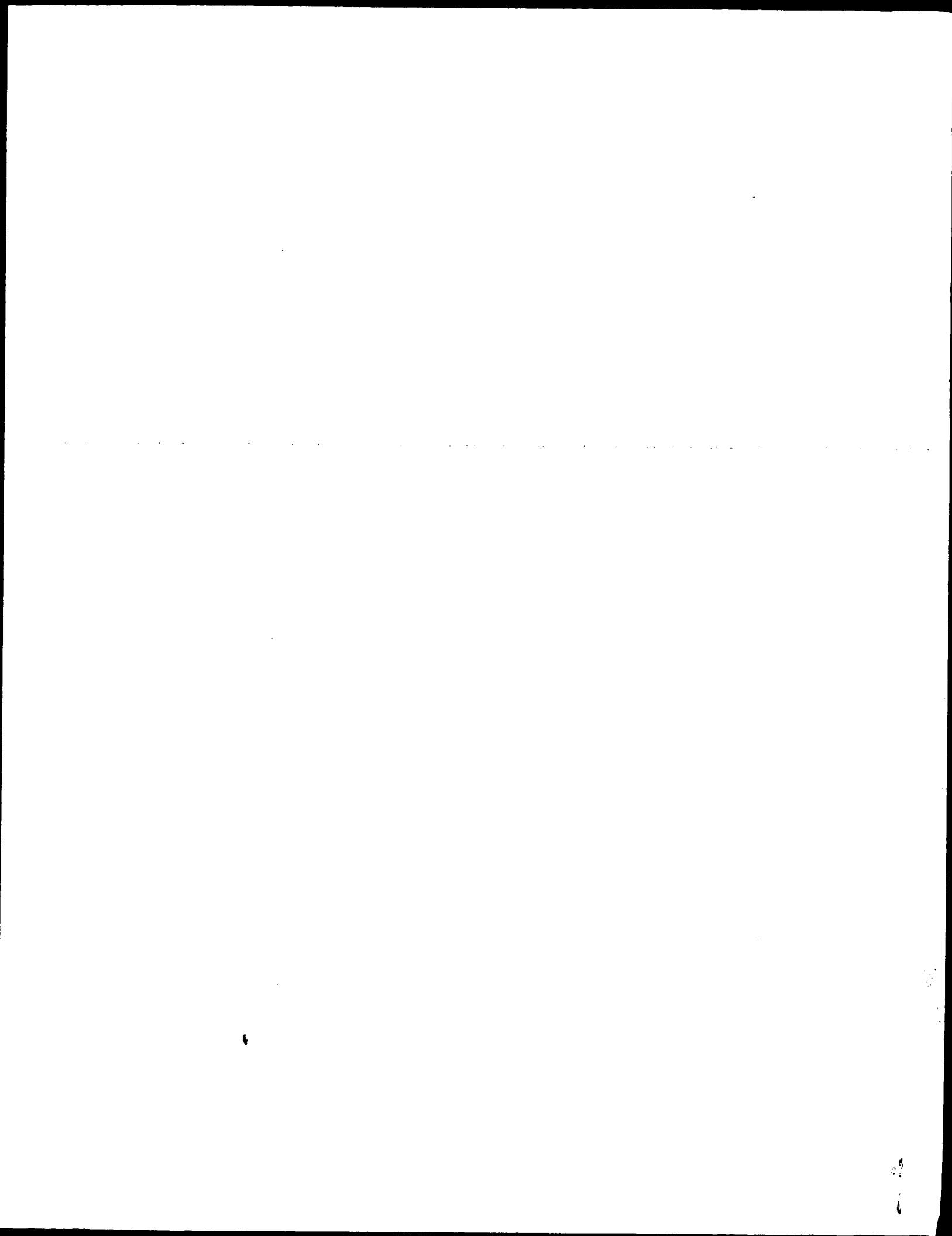
KW Neuron-specific transcription; promoter; mouse; Presenilin-1 gene;  
KW transgenic; laboratory model; Alzheimer's disease; ss.  
OS  
XX Mus sp.  
XX  
PN WO9910368-A1.  
XX  
PD 04-MAR-1999.  
XX  
PF 28-AUG-1998; 98WO-US17905.  
XX  
PR 29-AUG-1997; 97US-0920422.  
XX  
PA (UYDU-) UNIV DUKE.  
XX  
PI Mitsuda N, Rosas AD, Vitek PM;  
XX WPI; 1999-312401/26.  
DR  
PT Presenilin-1 gene promoter  
XX  
PS Claim 1; Page 51-81; 85pp; English.  
XX  
CC The invention provides a DNA molecule that directs neuron-specific  
CC transcription of a promoter segment in a mammalian cell. The promoter  
CC segment is from the mouse genomic Presenilin-1 gene. Transgenic non-human  
CC mammals containing a DNA expression cassette comprising the neuron-  
CC specific promoter are useful as laboratory models for studying the  
CC function of the Presenilin gene, and for studying the etiology of  
CC Alzheimer's disease.  
XX  
SQ Sequence 48974 BP; 12579 A; 10749 C; 11189 G; 14438 T; 19 other;  
  
Query Match 7.0%; Score 67.6; DB 20; Length 48974;  
Best Local Similarity 66.7%; Pred. No. 1.7e-09;  
Matches 112; Conservative 0; Mismatches 54; Indels 2; Gaps 1;  
  
Qy 287 gctacagccagcgggtctcaccctctcctgagcttgacccttaatacagtgctca 346  
Db 15137 gcatctaaagcagcgggtctcaccctctcctgagcttgacccttaatacagtgctca 15196  
  
Qy 347 tgcctgggtgncctcccaaccttaaaattttttgttgcctcacaactgtgatttt 406  
Db 15197 tctcgtgtggc--cccaatcataacatttttcttcttctcctcactgtaattct 15254  
  
Qy 407 gatactgttagaattgtaataataataatttgaagaagaggtttg 454  
Db 15255 gttgctgtta:gaatcataataataatcggatcgaggatctctg 15302  
  
RESULT 11  
Z23892  
ID Z23892 standard; DNA; 36901 BP.  
XX  
AC Z23892;  
XX  
DT 25-JAN-2000 (first entry)  
XX  
DE Murine LOBO genomic DNA fragment 2.  
XX  
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;  
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;  
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.  
XX  
OS Mus musculus.  
XX  
PN WO9950284-A2.  
XX  
PD 07-OCT-1999.  
XX  
PF 26-MAR-1999; 99WO-EP02055.  
XX











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:11:47 ; Search time 141.08 Seconds  
(without alignments)  
1029.145 Million cell updates/sec

Title: US-09-101-423A-5  
Perfect score: 960  
Sequence: 1 gaggggggtggcagatt.....tcaaccacccaaccacacat 960

Scoring table: IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	96	10.0	2830	2	US-08-604-333-3
C 2	96	10.0	2830	5	US-09-110-618-3
C 3	79.8	8.3	7208	5	US-09-166-186-107
C 4	62.6	6.5	16442	5	US-08-781-891-208
C 5	49.6	5.2	29604	5	US-08-781-891-207
C 6	48.6	5.1	5687	3	US-08-380-403A-3
C 7	48.6	5.1	5687	4	US-08-895-628-3
C 8	41.6	4.3	10614	2	US-08-135-511-35
C 9	41.6	4.3	10614	1	US-08-187-453-35
C 10	41	4.3	6845	3	US-08-380-403A-4
C 11	41	4.3	6845	4	US-08-895-628-4
C 12	39.8	4.1	26700	2	US-08-472-217-1
C 13	39.8	4.1	26700	3	US-08-488-199-5
C 14	39.8	4.1	26700	5	US-08-760-534A-1
C 15	37.6	3.9	4322	2	US-08-673-789-1
C 16	36.4	3.8	3162	4	US-08-449-645A-12
C 17	36.4	3.8	3162	4	US-08-702-367A-12
C 18	36.4	3.8	3162	6	PCT-US95-04681-12
C 19	35.2	3.7	278	1	US-08-248-474-106
C 20	35.2	3.7	278	5	US-08-756-849-106
C 21	35	3.6	3450	3	US-08-378-617A-9
C 22	34.8	3.6	387	5	US-09-157-177-135
C 23	33.6	3.5	7218	1	US-08-232-463-14
C 24	33.4	3.5	509	2	US-08-691-641-6
C 25	31	3.2	1157	1	US-08-383-756-4
C 26	31	3.2	1157	3	US-08-460-898-4

27	31	3.2	1157	4	US-08-440-845D-7
C 28	30.8	3.2	490	6	PCT-US95-08295-23
C 29	30.8	3.2	2623	5	US-08-804-439A-6
C 30	30.8	3.2	2623	5	US-08-720-429-6
C 31	30.8	3.2	72928	5	US-09-009-913-1
C 32	30.4	3.2	2624	1	US-08-232-144-3
C 33	30.4	3.2	4571	1	US-08-232-144-5
C 34	30.2	3.1	6152	2	US-08-557-139-1
C 35	30	3.1	19124	4	US-08-487-826B-13
C 36	29.8	3.1	232	3	US-08-434-557C-85
C 37	29.8	3.1	232	4	US-08-340-426D-85
C 38	29.8	3.1	232	4	US-08-450-673C-85
C 39	29.8	3.1	232	6	PCT-US95-17111A-85
C 40	29.8	3.1	565	4	US-08-731-722-7
C 41	29.8	3.1	934	4	US-08-731-722-6
C 42	29.8	3.1	934	4	US-08-731-722-6
C 43	29.8	3.1	1183	4	US-08-731-722-8
C 44	29.8	3.1	1183	4	US-08-731-722-8
C 45	29.8	3.1	1186	4	US-08-731-722-5

## ALIGNMENTS

RESULT 1  
US-08-604-333-3/c  
; Sequence 3, Application US/08604333  
; Patent No. 5776731  
; GENERAL INFORMATION:  
; APPLICANT: Parnet, Patricia et al.  
; TITLE OF INVENTION: Receptor Designated 2F1  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kathryn A. Anderson, Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple 7.1  
; SOFTWARE: Microsoft Word, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/604,333  
; FILING DATE: 21-FEB-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Anderson, Kathryn A.  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2619  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; TELEX: 756822  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2830 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHEetical: NO  
; ANTI-SENSE: NC  
; IMMEDIATE SOURCE:  
; CLONE: mu2F1  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 381..1994  
; NAME/KEY: mat\_peptide  
; LOCATION: 435..1991

FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 381..434  
US-08-604-333-3

Query Match 10.0%; Score 96; DB 2; Length 2830;  
Best Local Similarity 77.2%; Pred. No. 1.3e-19;  
Matches 142; Conservative 0; Mismatches 40; Indels 2; Gaps 2;

QY 289 ctacagccagcgggtctcaacccctctgctgctgaccccttaataacagtcgcctcatg 348  
Db 2440 CTCTAGCCAGCGGTCTCAACCTTCTTAATGCTTCAACCTTTAATACAGTCTTCACA 2381  
QY 349 ctctgggacccctcccaacccctaaattttttgtgtct-gttcataactgtgatttg 407  
Db 2380 CTCTGGTGAACCCCAACCATFAAACTTTTCGTGTGCTACTTGTGAAGTGTGG 2321  
QY 408 atactgttatgaattgtaataataat-aattttgagaagaagggttggccaaagggttga 466  
Db 2320 CTACAGTTATGAATCATAGCGTACATCTATTTGGAGATAAAGGTTTGCCTAAAGGGGCCA 2261  
QY 467 gaac 470  
Db 2260 CAAC 2257

RESULT 2  
US-09-110-618-3/c  
Sequence 3, Application US/09110618  
Patent No. 6050918  
GENERAL INFORMATION:  
APPLICANT: Parnet, Patricia et al.  
TITLE OF INVENTION: Receptor Designated 2F1  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kathryn A. Anderson, Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/110,618  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/604,333  
FILING DATE: 21-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Anderson, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2619  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2830 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: muzf1

FEATURE:  
NAME/KEY: CDS  
LOCATION: 381..1994  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 433..1991  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 381..434  
US-09-110-618-3

Query Match 10.0%; Score 96; DB 5; Length 2830;  
Best Local Similarity 77.2%; Pred. No. 1.3e-19;  
Matches 142; Conservative 0; Mismatches 40; Indels 2; Gaps 2;

QY 289 ctacagccagcgggtctcaacccctctgctgctgaccccttaataacagtcgcctcatg 348  
Db 2440 CTCTAGCCAGCGGTCTCAACCTTCTTAATGCTTCAACCTTTAATACAGTCTTCACA 2381  
QY 349 ctctgggacccctcccaacccctaaattttttgtgtct-gttcataactgtgatttg 407  
Db 2380 CTCTGGTGAACCCCAACCATFAAACTTTTCGTGTGCTACTTGTGAAGTGTGG 2321  
QY 408 atactgttatgaattgtaataataat-aattttgagaagaagggttggccaaagggttga 466  
Db 2320 CTACAGTTATGAATCATAGCGTACATCTATTTGGAGATAAAGGTTTGCCTAAAGGGGCCA 2261  
QY 467 gaac 470  
Db 2260 CAAC 2257

RESULT 3  
US-09-166-186-107  
Sequence 107, Application US/09166186A  
Patent No. 6080580  
GENERAL INFORMATION:  
APPLICANT: Baker, Brenda  
APPLICANT: Bennett, C. Frank  
APPLICANT: Butler, Madeline M.  
APPLICANT: Shanahan, William R.  
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDE MODULATION OF TNF-a EXPRESSION  
FILE REFERENCE: ISPH-0322  
CURRENT APPLICATION NUMBER: US/09/166,186A  
CURRENT FILING DATE: 1998-10-05  
NUMBER OF SEQ ID NOS: 250  
SEQ ID NO 107  
LENGTH: 7208  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (4527..4712,5225..5279,5457..5504,5799..6217)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (4371)..(4712)  
FEATURE:  
NAME/KEY: intron  
LOCATION: (4713)..(5224)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (5225)..(5279)  
FEATURE:  
NAME/KEY: intron  
LOCATION: (5280)..(5456)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (5457)..(5504)  
FEATURE:  
NAME/KEY: intron  
LOCATION: (5505)..(5798)  
FEATURE:

NAME/KEY: exon  
LOCATION: (5799)...(>6972)  
PUBLICATION INFORMATION:  
AUTHORS: Semon, D.  
AUTHORS: Kawashima, E.  
AUTHORS: Jongeneel, C.V.  
AUTHORS: Shakhov, A.N.  
AUTHORS: Nedospasov, S.A.  
TITLE: Nucleotide sequence of the murine TNF locus, including the  
TITLE: TNF-alpha (tumor necrosis factor) and TNF-beta (lymphotoxin) genes  
JOURNAL: Nucleic Acids Res.  
VOLUME: 15  
ISSUE: 21  
PAGES: 9083-9084  
DATE: 1987-11-11  
DATABASE ACCESSION NUMBER: Y00467 Genbank  
DATABASE ENTRY DATE: 1993-05-11  
US-09-166-186-107

Query Match 8.3%; Score 79.8; DB 5; Length 7208;  
Best Local Similarity 80.4%; Pred. No. 1.7e-14;  
Matches 131; Conservative 0; Mismatches 27; Indels 5; Gaps 3;

QY 301 gttctcaacctctcgtatgttgaccccttataacagtcctcagtcctcgtgacct 360  
|||||  
Db 126 gttctcaacctctcgtatgttgaccccttataacagtcctcagtcctcgtgacct 182  
QY 361 ccccaaccttataatttttgcgt-gtccataactgtgatttgatactgttatga 419  
|||||  
Db 183 ccccaaccttataatttttgcgt-gtccataactgtgatttgatactgttatga 242  
QY 420 attgtaataataaatt-ttgaagaagaaggtttgccaagg 461  
|||||  
Db 243 atcataagtataattttgttttaataagaggtttgccaagg 285

RESULT 4  
US-08-781-891-208  
; Sequence 208, Application US/08781891  
; Patent No. 6090620  
; GENERAL INFORMATION:  
; APPLICANT: Fu, Ying-Hui  
; APPLICANT: Yu, Chang-En  
; APPLICANT: Oshima, Junko  
; APPLICANT: Mulligan, John T.  
; APPLICANT: Schellenberg, Gerald D.  
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
; TITLE OF INVENTION: WERNER'S SYNDROME  
; NUMBER OF SEQUENCES: 209  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,891  
; FILING DATE: 27-DEC-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6090620tenburg Ph.D., Carol  
; REGISTRATION NUMBER: 39,317  
; REFERENCE/DOCKET NUMBER: 240052.419  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 208:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16442 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-781-891-208

Query Match 6.5%; Score 62.6; DB 5; Length 16442;  
Best Local Similarity 70.4%; Pred. No. 4.4e-09;  
Matches 112; Conservative 0; Mismatches 44; Indels 3; Gaps 2;  
QY 296 ccaggggttcacaccttcctcgtcgtcccttataacagtcctcagtcctcgtgacct 355  
|||||  
Db 5851 CCAGTGGTTCCTCAACCTTCCTG-TACTGTGACCCCTTATACAGTTCCTCGTGTGTAT 5909  
QY 356 gacctcccccacaccttataatttttggctgttcataacagtcgtgatttgatactgtt 415  
|||||  
Db 5910 GGCACCAACCATGACATTATTCCTTTGCTAG--TTCTTGACTGTAATTTCCACCGTT 5967  
QY 416 atgaattgtaataataataattttgagaagaagaggtttg 454  
|||||  
Db 5968 ATGAATTGTGATGTAATACTATCTGATATACAGGATGTTG 6006

RESULT 5  
US-08-781-891-207/c  
; Sequence 207, Application US/08781891  
; Patent No. 6090620  
; GENERAL INFORMATION:  
; APPLICANT: Fu, Ying-Hui  
; APPLICANT: Yu, Chang-En  
; APPLICANT: Oshima, Junko  
; APPLICANT: Mulligan, John T.  
; APPLICANT: Schellenberg, Gerald D.  
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
; TITLE OF INVENTION: WERNER'S SYNDROME  
; NUMBER OF SEQUENCES: 209  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,891  
; FILING DATE: 27-DEC-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6090620tenburg Ph.D., Carol  
; REGISTRATION NUMBER: 39,317  
; REFERENCE/DOCKET NUMBER: 240052.419  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 207:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 29604 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-781-891-207  
Query Match 5.2%; Score 49.6; DB 5; Length 29604;  
Best Local Similarity 65.2%; Pred. No. 5.3e-05;





COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/380,403A  
FILING DATE: 30-JAN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/325,909  
FILING DATE: 19-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-279712  
FILING DATE: 20-OCT-1994  
APPLICATION NUMBER: JP 6-139513  
FILING DATE: 30-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/128/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6645 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(904, 1015, 1356..1459, 1726..1883, 2009  
LOCATION: ..2618, 2890..3164, 4291..4509, 4598..4709, 4795  
LOCATION: ..4903, 5017..5117, 5200..5255, 5447..5525, 5598  
LOCATION: ..5741)  
US-08-380-403A-4

Query Match 4.3%; Score 41; DB 3; Length 6645;  
Best Local Similarity 62.8%; Pred. No. 0.011;  
Matches 81; Conservative 0; Mismatches 45; Indels 3; Gaps 1;  
QY 327 cccttaatacagtgctcgtcgtgacccctcccaacccctaaattattttgttg 386  
Db 386 CCCTTCCACAGCTCCACATGGTGTGTGA---CCCTGATCAAAAATATTTTCATGCG 442  
QY 387 ctgttcataactggtatttgatgactgttatgaattgtaataataattttgaagaaa 446  
Db 443 CACTTCATAACTGTAATTCTGCTACCTTAGGAATGTACCATAAACCACACGACACAG 502  
QY 447 gaggtttgc 455  
Db 503 GGTCTCTGC 511

RESULT 11  
US-08-895-628-4  
Sequence 4, Application US/08895628  
Patent No. 598585  
GENERAL INFORMATION:  
APPLICANT: MINATO, Negahiro  
APPLICANT: HATTORI, Masakazu  
APPLICANT: HIROSHI, Kubota  
APPLICANT: MASATSUGU, Maeda  
TITLE OF INVENTION: SPA-1 PROTEIN AND GENE CODING THEREFOR  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500

CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/895,628  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/380,403  
FILING DATE: 30-JAN-1995  
APPLICATION NUMBER: US 08/325,909  
FILING DATE: 19-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-279712  
FILING DATE: 20-OCT-1994  
APPLICATION NUMBER: JP 6-139513  
FILING DATE: 30-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/128/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6645 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(904, 1015, 1356..1459, 1726..1883, 2009  
LOCATION: ..2618, 2890..3164, 4291..4509, 4598..4709, 4795  
LOCATION: ..4903, 5017..5117, 5200..5255, 5447..5525, 5598  
LOCATION: ..5741)  
US-08-895-628-4

Query Match 4.3%; Score 41; DB 4; Length 6645;  
Best Local Similarity 62.8%; Pred. No. 0.011;  
Matches 81; Conservative 0; Mismatches 45; Indels 3; Gaps 1;  
QY 327 cccttaatacagtgctcgtcgtgacccctcccaacccctaaattattttgttg 386  
Db 386 CCCTTCCACAGCTCCACATGGTGTGTGA---CCCTGATCAAAAATATTTTCATGCG 442  
QY 387 ctgttcataactggtatttgatgactgttatgaattgtaataataattttgaagaaa 446  
Db 443 CACTTCATAACTGTAATTCTGCTACCTTAGGAATGTACCATAAACCACACGACACAG 502  
QY 447 gaggtttgc 455  
Db 503 GGTCTCTGC 511

RESULT 12  
US-08-472-217-1/c  
Sequence 1, Application US/08472217  
Patent No. 5726058  
GENERAL INFORMATION:  
APPLICANT: Alähen-Kurki, Leena  
APPLICANT: Auvinen, Petri  
APPLICANT: Jaakkola, Panu



APPLICANT: Jalkanen, Markku  
APPLICANT: Lepp, Sirpa  
APPLICANT: Mali, Markku  
APPLICANT: Viinen, Tapani  
APPLICANT: W ri, Anni  
TITLE OF INVENTION: Syndecan Stimulation of Cellular  
Differentiation  
TITLE OF INVENTION: Syndecan Stimulation of Cellular  
Differentiation  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.

STALE: D.C.S.A.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,217

FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/206,186  
 FILING DATE: 07-MAR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/988,427

ATTORNEY/AGENT INFORMATION:  
FILING DATE: 01-DEC-1992  
NAME: Cimbala, Michele A.  
REGISTRATION NUMBER: 33,851  
REFERENCE/STOCK NUMBER: 1102.0050003

TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 26700 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(4378..4443,

LOCATION: 23905..24039, 24039-24040, 24040-24041, 24041-24042, 24042-24043, 24043-24044, 24044-24045, 24045-24046, 24046-24047, 24047-24048, 24048-24049, 24049-24050, 24050-24051, 24051-24052, 24052-24053, 24053-24054, 24054-24055, 24055-24056, 24056-24057, 24057-24058, 24058-24059, 24059-24060, 24060-24061, 24061-24062, 24062-24063, 24063-24064, 24064-24065, 24065-24066, 24066-24067, 24067-24068, 24068-24069, 24069-24070, 24070-24071, 24071-24072, 24072-24073, 24073-24074, 24074-24075, 24075-24076, 24076-24077, 24077-24078, 24078-24079, 24079-24080, 24080-24081, 24081-24082, 24082-24083, 24083-24084, 24084-24085, 24085-24086, 24086-24087, 24087-24088, 24088-24089, 24089-24090, 24090-24091, 24091-24092, 24092-24093, 24093-24094, 24094-24095, 24095-24096, 24096-24097, 24097-24098, 24098-24099, 24099-24100, 24100-24101, 24101-24102, 24102-24103, 24103-24104, 24104-24105, 24105-24106, 24106-24107, 24107-24108, 24108-24109, 24109-24110, 24110-24111, 24111-24112, 24112-24113, 24113-24114, 24114-24115, 24115-24116, 24116-24117, 24117-24118, 24118-24119, 24119-24120, 24120-24121, 24121-24122, 24122-24123, 24123-24124, 24124-24125, 24125-24126, 24126-24127, 24127-24128, 24128-24129, 24129-24130, 24130-24131, 24131-24132, 24132-24133, 24133-24134, 24134-24135, 24135-24136, 24136-24137, 24137-24138, 24138-24139, 24139-24140, 24140-24141, 24141-24142, 24142-24143, 24143-24144, 24144-24145, 24145-24146, 24146-24147, 24147-24148, 24148-24149, 24149-24150, 24150-24151, 24151-24152, 24152-24153, 24153-24154, 24154-24155, 24155-24156, 24156-24157, 24157-24158, 24158-24159, 24159-24160, 24160-24161, 24161-24162, 24162-24163, 24163-24164, 24164-24165, 24165-24166, 24166-24167, 24167-24168, 24168-24169, 24169-24170, 24170-24171, 24171-24172, 24172-24173, 24173-24174, 24174-24175, 24175-24176, 24176-24177, 24177-24178, 24178-24179, 24179-24180, 24180-24181, 24181-24182, 24182-24183, 24183-24184, 24184-24185, 24185-24186, 24186-24187, 24187-24188, 24188-24189, 24189-24190, 24190-24191, 24191-24192, 24192-24193, 24193-24194, 24194-24195, 24195-24196, 24196-24197, 24197-24198, 24198-24199, 24199-24200, 24200-24201, 24201-24202, 24202-24203, 24203-24204, 24204-24205, 24205-24206, 24206-24207, 24207-24208, 24208-24209, 24209-24210, 24210-24211, 24211-24212, 24212-24213, 24213-24214, 24214-24215, 24215-24216, 24216-24217, 24217-24218, 24218-24219, 24219-24220, 24220-24221, 24221-24222, 24222-24223, 24223-24224, 24224-24225, 24225-24226, 24226-24227, 24227-24228, 24228-24229, 24229-24230, 24230-24231, 24231-24232, 24232-24233, 24233-24234, 24234-24235, 24235-24236, 24236-24237, 24237-24238, 24238-24239, 24239-24240, 24240-24241, 24241-24242, 24242-24243, 24243-24244, 24244-24245, 24245-24246, 24246-24247, 24247-24248, 24248-24249, 24249-24250, 24250-24251, 24251-24252, 24252-24253, 24253-24254, 24254-24255, 24255-24256, 24256-24257, 24257-24258, 24258-24259, 24259-24260, 24260-24261, 24261-24262, 24262-24263, 24263-24264, 24264-24265, 24265-24266, 24266-24267, 24267-24268, 24268-24269, 24269-24270, 24270-24271, 24271-24272, 24272-24273, 24273-24274, 24274-24275, 24275-24276, 24276-24277, 24277-24278, 24278-24279, 24279-24280, 24280-24281, 24281-24282, 24282-24283, 24283-24284, 24284-24285, 24285-24286, 24286-24287, 24287-24288, 24288-24289, 24289-24290, 24290-24291, 24291-24292, 24292-24293, 24293-24294, 24294-24295, 24295-24296, 24296-24297, 24297-24298, 24298-24299, 24299-24300, 24300-24301, 24301-24302, 24302-24303, 24303-24304, 24304-24305, 24305-24306, 24306-24307, 24307-24308, 24308-24309, 24309-24310, 24310-24311, 24311-24312, 24312-24313, 24313-24314, 24314-24315, 24315-24316, 24316-24317, 24317-24318, 24318-24319, 24319-24320, 24320-24321, 24321-24322, 24322-24323, 24323-24324, 24324-24325, 24325-24326, 24326-24327, 24327-24328, 24328-24329, 24329-24330, 24330-24331, 24331-24332, 24332-24333, 24333-24334, 24334-24335, 24335-24336, 24336-24337, 24337-24338, 24338-24339, 24339-24340, 24340-24341, 24341-24342, 24342-24343, 24343-24344, 24344-24345, 24345-24346, 24346-24347, 24347-24348, 24348-24349, 24349-24350, 24350-24351, 24351-24352, 24352-243

298 agcggttctcaaccttctgatgc

20102 AGTGGCTCTCAACCCCTCCTAACAC  
|||  
358 cctccccaaccttaaaattatttt  
|||  
20042 tttcttctgggtctt  
-----

20042 111C18811C1  
418 gaattgtaataataatttga  
111 111 111 111 111  
20002 GAATCATAATGTAATATCTAATA

ULT 13  
08-488-199-5/c  
Sequence 5, Application US/08488

1

RESULT 14  
US-08-760-534A-1/c  
; Sequence 1, Application US/08760534A  
; Patent No. 6017727  
; GENERAL INFORMATION:  
; APPLICANT: JALKANEN, MARKKU  
; APPLICANT: JAAKKOLA, PANU  
; APPLICANT: VIHINEN, TAPANI  
; TITLE OF INVENTION: SYNDECAN ENHANCER ELEMENT AND SYNDECAN  
; TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
; STREET: 1100 NEW YORK AVENUE, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: US  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/760,534A  
; FILING DATE: 02-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/206,186  
; FILING DATE: 07-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FI93/00514  
; FILING DATE: 01-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CIMBALA, MICHELE A.  
; REGISTRATION NUMBER: 33,851  
; REFERENCE/DOCKET NUMBER: 1708.0050004/MAC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26700 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(4378..4443, 22026..22106, 23001..23483,  
; LOCATION: 23905..24039, 24251..24418)  
US-08-760-534A-1

Query Match 4.1%; Score 39.8; DB 5; Length 26700;  
Best Local Similarity 61.7%; Pred. No. 0.048;  
Matches 92; Conservative 0; Mismatches 37; Indels 20; Gaps 1;  
QY 298 agcggttcacccctcctgctgctgaccccttaatacagtcctcctgctggtga 357  
Db 20102 AGTGGCCTCAACCCCTCTACACCGGACCCCTTAATACAGGTTCTCATGTCGGTGA 20043  
QY 358 cctcccaacccctaaataattttgtgtgttcataactggtgatttgatactgttat 417  
Db 20042 TTCTTGGTCTCT-----TTACATTTGTAATTTGCTACAGTTAT 20003  
QY 418 gaattgtaataataaattttgaagaaa 446  
Db 20002 GAATCAATATGTAATATCTAATATGTA 19974

RESULT 15

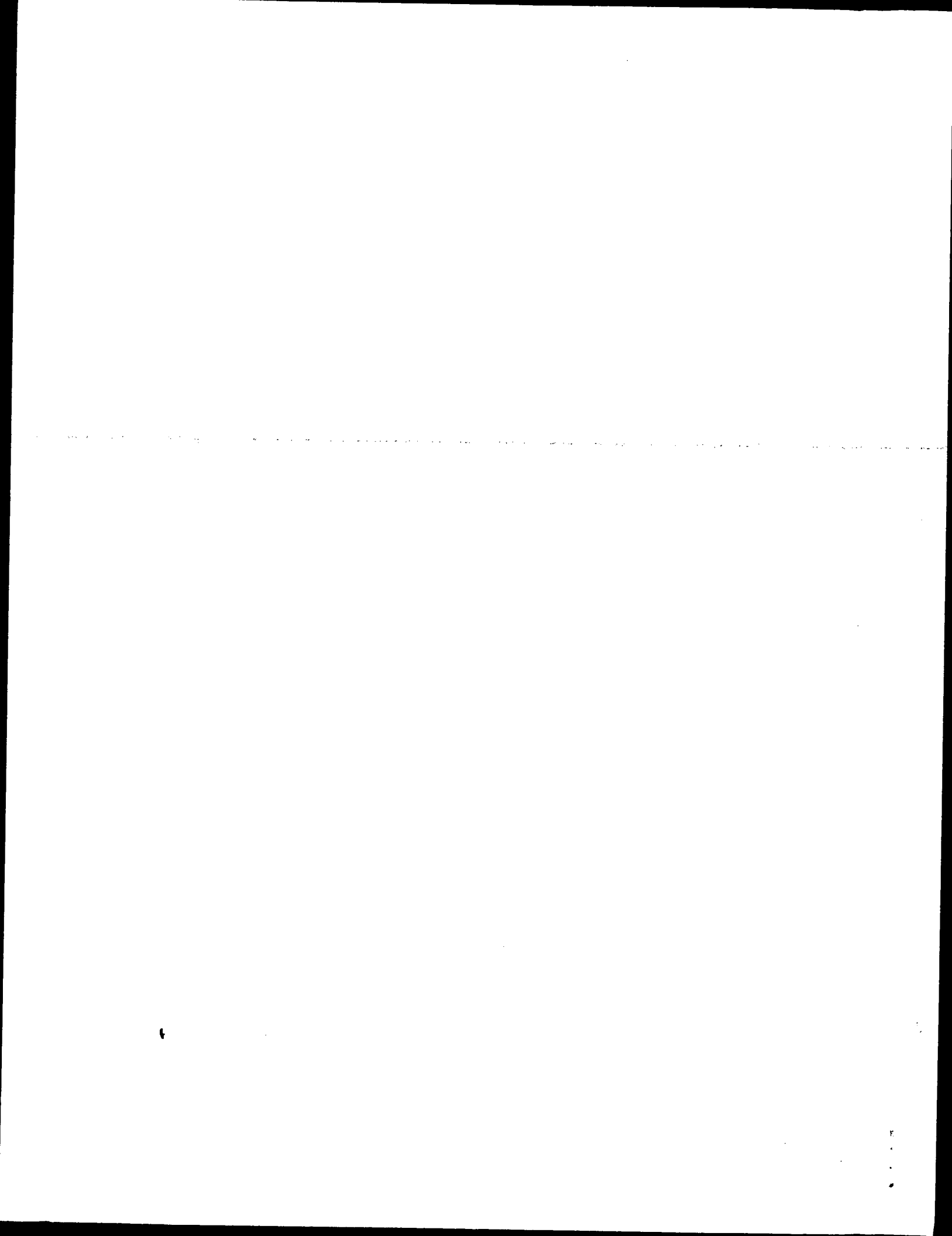
US-08-673-789-1  
; Sequence 1, Application US/08673789  
; Patent No. 5814479  
; GENERAL INFORMATION:  
; APPLICANT: ZHOOT, RENPING; SCHULZ, NICHOLAS,  
; APPLICANT: T.; KROMER, LAWRENCE, F.; VANDE WOUDE,  
; APPLICANT: GEORGE, F.  
; TITLE OF INVENTION: BSK RECEPTOR LIKE  
; TITLE OF INVENTION: TYROSINE KINASE AND LIGAND AND THEIR  
; TITLE OF INVENTION: USE IN DIAGNOSTIC AND THERAPEUTIC  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/673,789  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/177,812  
; FILING DATE: 04-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CAROL M. GRUPPI  
; REGISTRATION NUMBER: 37,341  
; REFERENCE/DOCKET NUMBER: 2026-4105  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 758-4800  
; TELEFAX: (312) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4322  
; TYPE: NUCLEOTIDE  
; STRANDEDNESS: DOUBLE  
; TOPOLOGY: UNKNOWN  
US-08-673-789-1

Query Match 3.9%; Score 37.6; DB 2; Length 4322;  
Best Local Similarity 50.6%; Pred. No. 0.094;  
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 310 ccttcctgctgcttcgaccccttaatacagtcctcctgctggtgaccccccacgt 369  
Db 539 CCCTCTGGAGTGCTCTCTCTGTGTGCGCGCTCGGGACCCCTTTGGCCAGCCCGCA 598  
QY 370 taaattattttgtgtgttcataactgatttgatactgttgaattgtaatat 429  
Db 599 ACGAAGTGAATTTGTGGATTGCGGCACGTGTCATGGGGACCTTGGATGGATTGCTTTC 658  
QY 430 aaataattttjaagaagaggtttcccaaggtttgagaactgctgtcttagccccagt 489  
Db 659 CAAAGAACGGGTGGGAAGAGATTGGTGAAGTTGATGAGAACTATGCCCCCATCCACAT 718

Search completed: November 4, 2000, 00:13:00  
Job time: 22184 sec

us-09-101-423a-5.rni

'Mon Nov 6 10:14:38 2000



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 3, 2000, 22:07:58 ; Search time 3273.53 Seconds  
(without alignments)  
1813.179 Million cell updates/sec

Title: US-09-101-423A-5  
Perfect score: 960  
Sequence: 1 gagggggggtgggcacagt.....tcaaccaaccaaccacat 960

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues 14379728  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_est3:  
4: gb\_est4:  
5: gb\_est5:  
6: gb\_est6:  
7: gb\_est7:  
8: gb\_est8:  
9: gb\_est9:  
10: gb\_est10:  
11: gb\_est11:  
12: gb\_est12:  
13: gb\_est13:  
14: gb\_est14:  
15: gb\_est15:  
16: gb\_est16:  
17: gb\_est17:  
18: gb\_est18:  
19: gb\_est19:  
20: gb\_est20:  
21: gb\_est21:  
22: gb\_est22:  
23: gb\_est23:  
24: gb\_est24:  
25: gb\_est25:  
26: gb\_est26:  
27: gb\_est27:  
28: gb\_est28:  
29: gb\_est29:  
30: gb\_est30:  
31: gb\_est31:  
32: gb\_est32:  
33: gb\_est33:  
34: gb\_est34:  
35: gb\_est35:  
36: gb\_est36:  
37: gb\_est37:  
38: gb\_est38:  
39: gb\_est39:  
40: gb\_est40:  
41: em\_estba:  
42: em\_estfun:  
43: em\_esthum1:

44: em\_esthum2:  
45: em\_esthum3:  
46: em\_esthum4:  
47: em\_esthum5:  
48: em\_esthum6:  
49: em\_esthum7:  
50: em\_esthum8:  
51: em\_esthum9:  
52: em\_esthum10:  
53: em\_esthum11:  
54: em\_esthum12:  
55: em\_esthum13:  
56: em\_esthum14:  
57: em\_esthum15:  
58: em\_esthum16:  
59: em\_esthum17:  
60: em\_esthum18:  
61: em\_esthum19:  
62: em\_esthum20:  
63: em\_estin1:  
64: em\_estin2:  
65: em\_estin3:  
66: em\_estin4:  
67: em\_estom:  
68: em\_estov1:  
69: em\_estov2:  
70: em\_estpl1:  
71: em\_estpl2:  
72: em\_estpl3:  
73: em\_estpl4:  
74: em\_estpl5:  
75: em\_estro1:  
76: em\_estro2:  
77: em\_estro3:  
78: em\_estro4:  
79: em\_estro5:  
80: em\_estro6:  
81: em\_estro7:  
82: em\_estro8:  
83: em\_estro9:  
84: em\_estro10:  
85: em\_estro11:  
86: em\_estro12:  
87: em\_estro13:  
88: gb\_gss1:  
89: gb\_gss2:  
90: gb\_gss3:  
91: gb\_gss4:  
92: em\_gss1:  
93: em\_gss2:  
94: em\_gss3:  
95: em\_gss4:  
96: gb\_gss5:  
97: gb\_gss6:  
98: gb\_gss7:  
99: gb\_gss8:  
100: gb\_gss9:  
101: em\_gss5:  
102: em\_gss6:  
103: em\_gss7:  
104: em\_gss8:  
105: em\_gss9:  
106: em\_gss10:  
107: em\_gss11:  
108: gb\_gss10:  
109: gb\_gss11:  
110: em\_gss12:  
111: gb\_gss12:  
112: gb\_gss13:  
113: gb\_gss14:  
114: gb\_gss15:  
115: gb\_gss16:  
116: gb\_gss17:



IMAGE:1531120 3' similar to gb:L35933 Mouse erythrocyte protein 4.2 (MOUSE); mRNA sequence.

ACCESSION BE134339  
VERSION BE134339.1 GI:8596839  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 443)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov

This clone is available royalty-free through LBNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:947220.

FEATURES  
source  
1..443  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1531120"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 134 a 93 c 99 g 117 t

Query Match 10.6%; Score 101.4; DB 33; Length 443;  
Best Local Similarity 83.2%; Pred. No. 7.9e-18;  
Matches 139; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

QY 299 gcggttctcaacctctgctgaccccttataacagtgccctatgctgtgac 358  
|||||  
Db 405 GTGGTCTCAACCTCTCTAATGCTGTA-TCCTAATACAGTCTCTCATGTTGGTGAC 347  
|||||

QY 359 ctcccaaccttaaaattttttgtgtgttctatacactgtgattgatactgttat 417  
|||||  
Db 346 TCCCCCAACCATAAATATTTTGTGTACTTCTAATCTAATTTTGTACTGTAG 287  
|||||

QY 418 gaattgataataaattttgaagaagaggtttgccagggttt 464  
|||||  
Db 286 GAATCGTAATGTAATATTTTGGTGATACAGGTTGCCAAGGGTT 240  
|||||

RESULT 3  
AA756105 458 bp mRNA EST 21-JAN-1998  
LOCUS v37904.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone  
DEFINITION IMAGE:1224630 5' similar to gb:L35933 Mouse erythrocyte protein 4.2 (MOUSE); mRNA sequence.

ACCESSION AA756105  
VERSION AA756105.1 GI:2803303  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 458)  
AUTHORS Marra,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

TITLE  
JOURNAL  
COMMENT

Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LBNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:650222

Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 399.

FEATURES  
source  
1..458  
/organism="Mus musculus"  
/strain="NIH/Swiss"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1224630"  
/clone\_lib="Stratagene mouse heart (#937316)"  
/sex="pooled"  
/tissue\_type="heart"  
/dev\_stage="13 day embryos"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: heart; Vector: pBluescript SK-; Site:1: ECORI; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

BASE COUNT 117 a 97 c 91 g 153 t

Query Match 10.4%; Score 99.6; DB 6; Length 458;  
Best Local Similarity 80.1%; Pred. No. 2.5e-17;  
Matches 117; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 288 ctacagcccgagcgtctcaacctctgctgaccccttataacagtgccctat 347  
|||||  
Db 246 CCTATAAGCAGCTGTTCTCAACCTCTCTGAAGTTGTGACCTTTAATACAGTCTCTGT 305  
|||||

QY 348 gctctggtgacctcccaaccttaaaattttgtgtgtgtgtgtgtgtgtgtgtg 407  
|||||  
Db 306 GTTGTGTGTGACCCCAACCAATAAATATTGTTGTGACTTCTAATGTAATTTTG 365  
|||||

QY 408 atactgttatgaattgtaataataat 433  
|||||  
Db 366 CTACGTGTTAAGTATCATATAATAAT 391  
|||||

RESULT 4  
AQ925253 580 bp DNA GSS 21-DEC-1999  
LOCUS RPCI-23-297K8.TV RPCI-23 Mus musculus genomic clone RPCI-23-297K8,  
DEFINITION DNA sequence.  
ACCESSION AQ925253  
VERSION AQ925253.1 GI:6614256  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 580)  
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akimret  
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroll,M., de Jong,P.  
and and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-23





## Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
or from Resea ch Genetics ([http://www.tigr.org/tigr/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tigr/bac_ends/mouse/bac_end_intro.html))  
plate: 328 row: A column: 19

FEATURES	Source
Seq primer: T7	
Class: BAC ends.	
Location/Qualifiers	
1. .562	
/organism="Mus musculus"	
/strain="C57BL/6J"	
/db_xref="taxon:10090"	
/clone_lib="RPCI-23-328A19"	
/clone_lib="RPCI-23"	
/sex="Female"	
/lab_host="DH10B"	
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: ECORI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the ECORI sites. The ligation products were transformed into ECORI sites." (BRL Life Technologies)."	
174 a	138 c
DH10B	electrocompetent cells (BRL Life Technologies).
	104 g
	145 t
	1 others
BAC COUNT	

BASE COUNT	174 a	138 c	104 g	145 t	1 others
ORIGIN					

Query Match	10.3%	Score 99	DB 113	Length 562
Best Local Similarity	70.4%	Pred. No.	3.9e-17	
Matches 178	Conservative	0	Mismatches 65	Indels 10
				Gaps 3

Qy	645	tttggctactgttggttccttttttcttccaacattttcta--ctcttttttccacctg	701
Db	168	TTTGTACTTTGTGGTTTTTCTCTCTCCACCCTCCCATCCCTTTCTCCACCCTT	227
Qy	702	tcgccccctaacactaaataagaagaaggggagcatagaggggaaagaaccccc	761
Db	228	TCACCCCCCAACACTAGATTAAGAGAGATAAAGGCCAGAGAGAAATGGAAGAGATCCC	287
Qy	762	tgaataacgttcagtagttggtccaaaggggggtgacatatgttgcatttagaccacatctctg	821
Db	288	TGAATAAAAGTCAGGGGCTGGAAA-----GTGGATTGAGCTTATCATCTAGACTACTTCCTG	341
Qy	822	gtgattaaggggagccaagttccttggggccaagtttgatctcttcgttgaacgatactcaa	881
Db	342	TGATTATGGGACACT-GAGTTCCTTGGGACAAGTTTGATCTTTCGTCTCAGGATATCTAG	400
Qy	882	ttctctctccctg	894
Db	401	TTTCTCTTTGTG	413

RESULT	7
AZ276484	
LOCUS	AZ276484 568 bp DNA
DEFINITION	RPCI-23-10707-TV RPCI-23 Mus musculus genomic clone RPCI-23-10707, GSS
ACCESSION	AZ276484
VERSION	AZ276484
KEYWORDS	DNA sequence.
SOURCE	AZ276484.1 GI:9495386
ORGANISM	GSS.
	house mouse,
	Mus musculus

Eukaryotes: Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 568)  
B. Levis M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-23

Unpublished (1993)  
Other GSSs: RPCI-23-10707.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel.: 301 838 0200  
Fax: 301 838 0208  
Email: s.zhao@igrr.org

Clones are derived from the mouse BAC library RPC1-23. For BAC library availability, please contact Pieter de Jong ([pieter@dejong.med.buffalo.edu](mailto:pieter@dejong.med.buffalo.edu)). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end page: [http://www.tlgr.org/cdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tlgr.org/cdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 107 row: 0 column: 7  
Seq primer: T7  
Class: EAC ends.

```

Location/Qualifiers
1. 568
./organism="Mus musculus"
./strain="C57BL/6J"
./db_xref="taxon:10090"
./clone="RPCI-23-10707"
./clone_lib="RPCI-23"
./sex="Female"
./lab_host="DH10B"
./note="Organ: Kidney/Brain; Vector: pBACE3 6; Site:1:
EcOri; Site:2: EcOri; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcOri and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcOri sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies).
178 a 139 c 104 g 147 t

```

BASE COUNT	178 a	139 c	104 g	147 t	ORIGIN
Query Match	10.3%	Score 98.8	DB 117	Length 568	
Best Local Similarity	65.9%	Pred. No. 4.5e-17			
Matches 179	Conservative	0	Mismatches 67	Indels 10	Gaps
QY	642	gagtttgctacttgtagtgggtcccttttcttcaaacctttttcta---ctctttttccacc	698		
Db	154	GATCTTTTGTACTTGTGTTTTACTCTCTCTCCACCCTTTCCCATCCCTTTTCTTCACC	213		
QY	699	ctgtcgccccctctaacactaataagaagaaggggagcatagagggaagaagaac	758		
Db	214	CTTCAACCCCCCAACACTAGATAGAGAGATAAAGGCCAGAGAGAATATGGAAGAGAT	273		
QY	759	ccttgaataaacctagttagtttgggcaaaagggggtgacatatgttgttcattagaccacatc	818		
Db	274	CCTGTGNTAAAGTCTAGGGGGCTGNA--GTTGGATGACGTTATCCTAGACTACTTTC	327		
QY	819	ctggttgattaaaggaggtcaagtttcccttggggcaagtttgattcttttgtaaacgatactc	878		
Db	328	CTGTGTATTAGTAGGACACT--GAGTTCCTTGGGACAAGTTTGATCTTTGCTGTCACATATC	386		

RESULT	8			
AZ029492				
LOCUS				
AZ029492				
465 bp				
DNA				
GSS				
25-FEB-2000				



Qy 882 ttctctccctg 894  
 ||||| ||  
 Db 398 TTTCTCTTTGTG 410

RESULT 10  
 AW489274/c

DEFINITION UI-M-BH3-ata-a-11-0-UI.s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
 UI-M-BH3-ata-a-11-0-UI 3', mRNA sequence.

ACCESSION AW489274  
 VERSION AW489274.1 GI:7059544

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE 1 (bases 1 to 446)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene  
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 9704477

COMMENT Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: MEST@mail.nih.gov

Oligo-OT track not found, Not 1 site shown in beginning of sequence  
 is likely internal to the message. cDNA Library Preparation: M.B.  
 Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENETICS. It should be noted that Bento Soares  
 is generating a small number of additional specialized  
 non-redundant arrays of BMAP cDNAs whose availability will be  
 considered under appropriate and limited collaborative arrangements.  
 The following repetitive elements were found in this cDNA sequence:  
 143-378, >URRIA=DNA/MERI\_type

Seq primer: M13 Forward

POLYA-No.

FEATURES Location/Qualifiers

source

1..446

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="NIH\_BMAP\_M\_S4"

/dev\_stage="27-32 days"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT73B-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not 1; Site\_2: Eco RI; The  
 NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
 ultimately derived from a mixture of individually tagged  
 normalized libraries from ten regions of the mouse brain  
 (cerebellum, brain stems, olfactory bulbs, hypothalamus,  
 cortex, amygdala, basal ganglia, pineal gland, striatum,  
 hippocampus) after a series of subtractions to reduce the  
 representation of cDNAs from which ESTs had already been  
 generated. The following serially subtracted libraries  
 were generated in this process: NIH\_BMAP\_M\_S4,  
 NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
 NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
 (NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
 cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
 NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
 was used as a driver in a hybridization with a pool of  
 the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
 libraries in the form of single-stranded circles. The  
 remaining single-stranded circles (subtracted library)  
 was purified by hydroxyapatite column chromatography,  
 converted to double-stranded circles and electroporated

into DH10B bacteria (Life Technologies) to generate the  
 NIH\_BMAP\_M\_S4 library. This procedure has been previously  
 described (Bonaldo, Lennon and Soares, Genome Research  
 6:791-806, 1996)  
 TAG\_LIB=NIH\_BMAP\_M\_S4  
 TAG\_TISSUE=hippocampus  
 TAG\_SEQ=TTCGA

BASE COUNT 114 a 111 g 130 t

ORIGIN

Query Match

Best Local Similarity 9.9%; Score 95.2; DB 22; Length 446;

Matches 124; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 293 agccacgggtctcaacctctgctgctgagcccttaataacagctcctcgtct 352

DB 377 AGGACAGTGTCTCCCAACCTTCTTAATGCTGGACCCCTTAATACAGTCTCCACACTGT 318

QY 353 ggtgacctcccaaccttaaaattttgtgtgtctgctcataaactgtgattttgatact 412

DB 317 AGTGACCCCAACCAACCAATTAATTAATCTGCTGCTACTTATCATTTGCTGCTACT 258

QY 413 gttatgaattgtaataataaattttgagaaaggggttgcacagggttt 464

DB 257 GTTACAAGTTGTAATGTAATAATCTGATATGCAGAAATAGCTGGCATGTGATT 206

RESULT 11

AZ038764/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other\_GSSs: RPCI-23-328F6.TV

Unpublished (1999)

Mouse BAC End Sequences from Library RPCI-23

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301' 838 0200

Fax: 301' 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html

Plate: 328 row: F column: 6

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..743

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="RPCI-23-328F6"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:

ECORI; Site\_2: EcorI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcorI and EcorI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcorI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT 236 a 128 c 177 g 201 t  
ORIGIN

Query Match 9.8%; Score 94.2; DB 114; Length 743;  
Best Local Similarity 77.0%; Pred. No. 9.3e-16;  
Matches 114; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 286 tgctacagccagcggcttcaacctctgatgcttgacaccttttaatacagtcgc 345  
DB 740 TGCTCCAGCAGTAGTGTTCTACCCCTCCAAATGCTGGACCCCTTAATACAGTCTCTC 681

QY 346 atgctctggtgacctcccccacaccttaaaattattttgtgtgtctacataactgtgatt 405  
DB 680 GTGTTATAGTACCCCAACCATANAATATTATTGTGTGTACATCATACTGTAATT 621

QY 406 tgatactgttatgaattgtaataataat 433  
DB 620 TGTACTGTATGAATGTAATGTAAT 593

RESULT 12  
AZ082208/c 627 bp DNA GSS 08-MAY-2000  
LOCUS  
DEFINITION RPCI-23-20B21.TJ RPCI-23 Mus musculus genomic clone RPCI-23-20B21, DNA sequence.

ACCESSION AZ082208  
VERSION AZ082208.1 GI:7724036  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 627)  
Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroi, M., de Jong, P. and and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other\_GSSs: RPCI-23-20B21.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from ResGen ([http://www.tigr.org/tdb/bac/ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html)). BAC end page: Plate: 20 row: B column: 21  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers  
1. .627  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-20B21"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;  
ECORI; Site\_2: EcorI; Female C57BL/6J mouse kidney and/or

FEATURES  
source

brain genomic DNA was isolated and partially digested with a combination of EcorI and EcorI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcorI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT 192 a 121 c 145 g 169 t  
ORIGIN

Query Match 9.7%; Score 93.4; DB 114; Length 627;  
Best Local Similarity 64.5%; Pred. No. 1.5e-15;  
Matches 211; Conservative 0; Mismatches 96; Indels 20; Gaps 4;

QY 636 tagctgagcttggttactgtgtgttcccttttttcaaaccttttctactctttttcc 695  
DB 340 TAGCTGTCTTGTACTTGTGGTCTTCTTCTTCACCCCTTCTCTACCCCTTTTC 281

QY 696 ac---cctg---cggcccccctaacactaaataagaagaagaggagcatagaggga 751  
DB 280 ATTATCTCTTCAATGCCCTTACACTAGATAGGAGAGAGAGAGAGAGAGAGGA 221

QY 752 aagaacacccctgaataacgtcagtagtggcaaaagggggtgacatattgtgtcattaga 811  
DB 220 AAGAGATCTCTGAATAAGTCAGGATCAGAAAGGGGTGACA-----TATTGTTAGA 168

QY 812 ccacatctctgtgattaaagggsgtcaagtctcttgggcaagtttgatcttctgtgtaa 871  
DB 167 TGACTTCTGTCTGATT---GGGGTTGAGTTTCTTGAGGCAAAATTTGATCTTTACTGTCA 111

QY 872 cgatatctaatctctctctgtcttctgtcttctgttggaacagacttgataccccc 931  
DB 110 AGATATCTAATTTCTTC-----CTGTTTCTTCTTTGACATGATTACTTACAAACTGT 57

QY 932 atggtaccattcaacacacacacacac 958  
DB 56 GGCCAAACAACCAACTCACCACC 30

RESULT 13  
B03243/c 541 bp DNA GSS 13-JUL-1996  
LOCUS  
DEFINITION CSRL-173D6-u CSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone CSRL-173D6, DNA sequence.

ACCESSION B03243  
VERSION B03243  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 541)  
Evans, G.A., Burbee, D., Davies, C., Hahner, L., Oliver, T., Gilbert, M., Jones, D., Ward, T., Gillilan, E., Schagemann, J., Probst, S., Harris, J., DeFord, J., McFarland, J., Burzinski, K., Khan, M., Kupfer, K. and Garner, H.R.

TITLE Genomic Sequence Sampled Map of Chromosome 11  
JOURNAL Unpublished (1996)  
COMMENT Contact: Evans GA, Shane Probst  
McDermott Center for Human Growth and Development  
University of Texas Southwestern Medical Center At Dallas  
5323 Harry Hines Blvd, Dallas TX 75235-8591  
Tel: 214-648-1600  
Fax: 214-648-1666  
Email: govans@utsw.swmed.edu, shane@mcdermott.swmed.edu

PCR PRIMERS  
FORWARD: TTTCACGAACTTTATATCC  
BACKWARD: ACAGATCAAGTCATGTCAC  
Seq primer: T7  
Class: cosmid ends  
High quality sequence stop: 541.  
Location/Qualifiers  
1. .541  
/organism="Homo sapiens"



Search completed: November 3, 2000, 22:08:03  
Job time: 26217 sec

DNA sequence.  
ACCESSION AZ068346  
VERSION AZ068346.1 GI:7359598  
KEYWORDS GSS.  
SOURCE house mouse  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 379)  
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.  
TITLE Mouse BAC End Sequences from Library RPCI-23  
JOURNAL Unpublished (1999)  
COMMENT Other GSSs: RPCI-23-390B4.TJ  
Contact: Shaving Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: [http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)  
Plate: 390 row: B column: 4  
Seq primer: T7  
Class: BAC ends.

## FEATURES

Location/Qualifiers  
1..379  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-390B4"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1: EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 105 a 88 c 69 g 117 t  
ORIGIN

Query Match 9.6%; Score 92.4; DB 114; Length 379;  
Best Local Similarity 69.2%; Pred. No. 2.6e-15;  
Matches 155; Conservative 0; Mismatches 66; Indels 3; Gaps 2;  
Qy 297 cagcggtctcaacgtctcgtatgcttcgtaccccttataacagtgccatgctctggtg 356  
Db 14 CAATGATCTTCAACCTCTTAACTGTCGCCACCTTTAATACAGTTCCTCATGTGTG 73  
Qy 357 accctcccaaccttaaatattttgtgtgttcataactgtgtttgatactgtt 415  
Db 74 ACCCT--CCAGCCAAAATAATTTAACTGCTATTCATTAACGTGTAATTTGCTGCTATT 131  
Qy 416 atgaattgataataataattttgaagaaggtttgccaaggtttgagaaactgctg 475  
Db 132 ATGAATTTCTAATAATAATATTTTGGAGATAGAGGTGTGCCAAAAGTCATAACCCCTCAG 191  
Qy 476 tictagccccacgtggatgttttcgtcatttgggtttttat 519  
Db 192 GATGAGAACTGATAAATCTGCCCTGCTCCGGTTGGGGCTTGTTT 235

Mon Nov 6 10:14:39 2000

us-09-101-423a-5.rst

---

Page 11

.....



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 04:20:02 ; Search time 7111.01 Seconds  
(without alignments)  
669.478 Million cell updates/sec

Title: US-09-101-423A-6  
Perfect score: 1090  
Sequence: 1 ttgtctctgtgttacttgt.....cctagcacagtcacccaca 1090

Scoring table: OLIGO/NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1033670 seqs, 2183789903 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pri.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: em\_fun.\*  
13: em\_hum1.\*  
14: em\_hum2.\*  
15: em\_in.\*  
16: em\_om.\*  
17: em\_or.\*  
18: em\_ov.\*  
19: em\_pat.\*  
20: em\_ph.\*  
21: em\_pl1.\*  
22: em\_ro.\*  
23: em\_sts.\*  
24: em\_sy.\*  
25: em\_un.\*  
26: em\_v1.\*  
27: gb\_htg1.\*  
28: gb\_htg2.\*  
29: gb\_in1.\*  
30: gb\_in2.\*  
31: em\_ba1.\*  
32: em\_ba2.\*  
33: em\_hum3.\*  
34: em\_hum4.\*  
35: gb\_pr4.\*  
36: gb\_htg3.\*  
37: gb\_htg4.\*  
38: gb\_htg5.\*  
39: gb\_htg6.\*  
40: gb\_htg7.\*  
41: em\_htg1.\*  
42: em\_htg2.\*  
43: em\_htg3.\*

44: em\_hum5.\*  
45: gb\_pl3.\*  
46: gb\_pr5.\*  
47: gb\_htg8.\*  
48: gb\_htg9.\*  
49: gb\_htg10.\*  
50: gb\_htg11.\*  
51: gb\_htg12.\*  
52: gb\_htg13.\*  
53: gb\_htg14.\*  
54: gb\_in3.\*  
55: gb\_htg15.\*  
56: gb\_htg16.\*  
57: gb\_htg17.\*  
58: em\_htg4.\*  
59: em\_htg5.\*  
60: em\_htg6.\*  
61: em\_htg7.\*  
62: em\_hum6.\*  
63: gb\_htg18.\*  
64: gb\_htg19.\*  
65: gb\_ba3.\*  
66: em\_htg8.\*  
67: em\_htg9.\*  
68: em\_htg10.\*  
69: em\_htg11.\*  
70: em\_htg12.\*  
71: em\_htg13.\*  
72: em\_htg14.\*  
73: em\_htg15.\*  
74: em\_htg16.\*  
75: em\_htg17.\*  
76: em\_htg18.\*  
77: em\_htg19.\*  
78: em\_htg20.\*  
79: em\_htg21.\*  
80: em\_htg22.\*  
81: em\_htg23.\*  
82: gb\_pr6.\*  
83: gb\_pr7.\*  
84: gb\_htg20.\*  
85: gb\_htg21.\*  
86: gb\_htg22.\*  
87: gb\_htg23.\*  
88: gb\_ro.\*  
89: gb\_sts1.\*  
90: gb\_sts2.\*  
91: gb\_sy.\*  
92: gb\_un.\*  
93: gb\_v11.\*  
94: gb\_v12.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1090	100.0	1090	5 A64063	A64063 Sequence 6
C 2	28	2.6	948	88 AF000201	AF000201 Rattus sp
C 3	28	2.6	1102	88 AF000205	AF000205 Rattus sp
C 4	28	2.6	1113	88 RN38NLV11	U87604 Rattus norv
C 5	28	2.6	1117	88 AF000202	AF000202 Rattus sp
C 6	28	2.6	1127	88 AF000200	AF000200 Rattus sp
C 7	28	2.6	1127	88 AF000203	AF000203 Rattus sp
C 8	28	2.6	1297	88 AF000204	AF000204 Rattus sp
C 9	28	2.6	1297	88 RN11MLV11	U87600 Rattus norv
C 10	28	2.6	1306	88 AF000199	AF000199 Rattus sp
C 11	28	2.6	1429	88 RN8MLV11	U87598 Rattus norv
C 12	28	2.6	1826	88 RN1RT5F	X61297 R.norvegicu

C 13	28	2.6	1857	88	RNLB6	X07686	Rat L1rn B6
C 14	28	2.6	1986	88	RNL14MLV1	U87602	Rattus norv
C 15	28	2.6	2225	88	AF058787	AF058787	Rattus norv
C 16	28	2.6	3047	88	RATEFLAA	M95791	Rat transcr
C 17	28	2.6	6335	88	RNLINED	X53581	R.norvegicu
C 18	28	2.6	8048	88	RATLN3A	M13100	Rat long in
C 19	28	2.6	16257	88	RNLVAX	X59496	Rat genes f
C 20	27	2.5	12288	88	RNLB7	X07687	Rat L1rn B7
C 21	27	2.5	13222	88	RATCYPAA2	M57719	Rat cytochr
C 22	26	2.4	6369	88	RATCRYGF	M19357	Rat gamma-F
C 23	25	2.3	4333	89	AU026950	AU026950	Rattus no
C 24	23	2.1	170272	56	AC069494	AC069494	Homo sapi
C 25	22	2.0	5577	88	RNL13	X03095	Rat LINE1 s
C 26	22	2.0	99886	11	AL135841	AL135841	Human DNA
C 27	22	2.0	100000	35	AP000131	AP000131	Homo sapi
C 28	22	2.0	100000	35	AP0000209	AP0000209	Homo sapi
C 29	22	2.0	146847	35	AP0000248	AP0000248	Homo sapi
C 30	22	2.0	149526	46	HS411B6	Z84470	Human DNA s
C 31	22	2.0	458305	48	AC023075	AC023075	Homo sapi
C 32	22	2.0	174256	48	AC0232572	AC0232572	Homo sapi
C 33	22	2.0	177311	9	AC004385	AC004385	Human Chr
C 34	22	2.0	184035	35	CNS01R1M	AL163842	Human chr
C 35	22	2.0	184740	35	AC011206	AC011206	Homo sapi
C 36	22	2.0	185570	63	AL138834	AL138834	Homo sapi
C 37	22	2.0	210480	84	AL355614	AL355614	Homo sapi
C 38	22	2.0	340000	35	AP001710	AP001710	Homo sapi
C 39	21	1.9	384	89	AU024858	AU024858	Rattus no
C 40	21	1.9	2946	88	S51970	S51970	CYP2B2-cyto
C 41	21	1.9	57056	86	AP000637	AP000637	Homo sapi
C 42	21	1.9	144563	10	AC006204	AC006204	Homo sapi
C 43	21	1.9	183918	50	AC025896	AC025896	Homo sapi
C 44	20	1.8	501	82	HSJ33319	AJ223319	Homo sapi
C 45	20	1.8	527	88	RNO223831	AJ223831	Rattus no

## ALIGNMENTS

RESULT	1
A64063	1090 bp DNA
LOCUS	
DEFINITION	Sequence 6 from Patent WO9725443.
ACCESSION	A64063
VERSION	A64063.1 GI:3717537
KEYWORDS	. unidentifed.
SOURCE	unidentified
ORGANISM	unclassified
REFERENCE	1. (bases 1 to 1090)
AUTHORS	Rudland,P.S. and Barraclough,B.R.
TITLE	METASTASIS INDUCING DNA'S
JOURNAL	Patent: WO 9725443-A 17-JUL-1997;
	UNIV LIVERPOOL (GB)
FEATURES	Location/Qualifiers
source	1..1090
	/organism="unidentified"
	/db_xref="taxon:32644"
BASE COUNT	224 a 277 c 275 g 314 t
ORIGIN	
Query Match	100.0%; Score 1090; DB 5; Length 1090;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1090; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ttgtctcgggtgactgttttccattctgcacagtggtttgaccttcatacgctg 60
Ddb	1 TTGTCCTCGGTACTGTGTTCCCAATTCTGCACAGTGGTTGACCTTCATACGCTG 60
QY	61 tgtctcaggagcgtgtagacctatttctcgttttcttcagccagtaccaggaacaga 120
Ddb	61 TGTGTTCAGGAGTGCTGTAGACCTATTTTCCTGTTTTCTTTTCAGCCAGTTACAGGAACAGA 120

```

binding protein gene, partial cds.
ACCESSION AF000201
VERSION AF000201.1 GI:2599310
KEYWORDS
SOURCE
ORGANISM
Rattus sp. T-612.
Rattus sp. T-612.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 948)
Cabot,E.L., Angeletti,B., Usdin,K. and Furano,A.V.
Rapid evolution of a young L1 (LINE-1) clade in recently speciated
Rattus taxa
JOURNAL J. Mol. Evol. 45 (4), 412-423 (1997)
MEDLINE 97472947
REFERENCE
2 (bases 1 to 948)
Cabot,E.L. and Furano,A.V.
Direct Submission
JOURNAL Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA
FEATURES
source
1..948
Location/Qualifiers
/organism="Rattus sp. T-612"
/isolate="T-612"
/db_xref="taxon:68896"
/note="from the collection of Institut des Sciences de
1'Evolution, Montpellier II; collected from the island of
Ceram, Indonesia and identified by Djoko Iskandar as
Rattus moluccarius; cranial morphology suggests similarity
to Rattus rattus satarae"
repeat_region 1..>948
/note="mlv12-tm5"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type="dispersed"
1..564
5'UTR
CDS
565..->948
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein"
/protein_id="AAC53496.1"
/db_xref="GI:2599311"
/translation="MARGKRRNPSSNRNDCMPPSPNPTKTNMEYPTPEKQDLVSK
SYLIMLEDFKDMTLRETOEILNKQVEAYREWQSKLKEFQENTIKQKELKMEIE
AIKKHMETTLDIENQKRGQAVDS"
BASE COUNT 355 a 234 c 225 g 134 t
ORIGIN
Query Match 2.6%; Score 28; DB 88; Length 948;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 gctgtgtgtcaggagtgtgttagacct 83
|||||
Db 503 GCCTGTGTGCAGGAGTGTGTAGACCT 476
|||||
RESULT 3
AF000205/c
LOCUS AF000205 1102 bp DNA ROD 11-DEC-1997
DEFINITION Rattus sp. T-612 retrotransposon mlv12-rml7 5'UTR and putative
binding protein gene, partial cds.
ACCESSION AF000205
VERSION AF000205.1 GI:2599318
KEYWORDS
SOURCE
ORGANISM
Rattus sp. T-612.
Rattus sp. T-612.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1102)
Cabot,E.L., Angeletti,B., Usdin,K. and Furano,A.V.
Rapid evolution of a young L1 (LINE-1) clade in recently speciated
Rattus taxa
JOURNAL J. Mol. Evol. 45 (4), 412-423 (1997)
MEDLINE 97472947
REFERENCE
2 (bases 1 to 1102)
Cabot,E.L. and Furano,A.V.
Direct Submission
JOURNAL Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA
FEATURES
source
1..1102
Location/Qualifiers
/organism="Rattus sp. T-612"
/isolate="T-612"
/db_xref="taxon:68896"
/note="from the collection of Institut des Sciences de
1'Evolution, Montpellier II; collected from the island of
Ceram, Indonesia and identified by Djoko Iskandar as
Rattus moluccarius; cranial morphology suggests similarity
to Rattus rattus satarae"
repeat_region 1..>1102
/note="mlv12-rml7"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type="dispersed"
1..718
5'UTR
CDS
719..->1102
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein"
/protein_id="AAC53500.1"
/db_xref="GI:2599319"
/translation="MARGKRRNPSSNRNDCMPPSPNPTKTNMEYPTPEKQDLVSK
SYLIMLEDFKDMTLRETOEILNKQVEAYREWQSKLKEFQENTIKQKELKMEIE
AIKKHMETTLDIENQKRGQAVDS"
BASE COUNT 385 a 293 c 271 g 153 t
ORIGIN
Query Match 2.6%; Score 28; DB 88; Length 1102;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 gctgtgtgtcaggagtgtgttagacct 83
|||||
Db 657 GCCTGTGTGCAGGAGTGTGTAGACCT 630
|||||
RESULT 4
RN38NLV1/c
LOCUS RN38NLV1 1113 bp DNA ROD 26-APR-1997
DEFINITION Rattus norvegicus L1 retrotransposon mlv12-rn38, 5'UTR and putative
RNA binding protein 1 gene, partial cds.
ACCESSION U87604
VERSION U87604.1 GI:2047334
KEYWORDS
SEGMENT 1 of 2
SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1113)
Hayward,B.E., Zavanelli,M. and Furano,A.V.
Recombination creates novel L1 (LINE 1) elements in Rattus
norvegicus
Genetics (1997) In press
2 (bases 1 to 1113)
Hayward,B.E., Good,B., Cabot,E. and Furano,A.V.
Direct Submission
JOURNAL Submitted (29-JAN-1997) LMCB/NIDDK, NIH, 8/203, 8 Center Dr. MSC
0830, Bethesda, MD 20892-0830, USA
FEATURES
source
1..1113
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"

```

```
/note="L1mlvrn38b"
5'UTR
repeat_region
<1..729
1..>1113
/note="mlv12-rn38"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type=dispersed
730..>1113
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein 1"
/protein_id="AAB53049.1"
/db_xref="GI:2047337"
/translation="MARGKRNPNSNRNODCMPSSEPNSTKTNNMYPNTPEKODLVSK
SYLIMLEDFKDMNTLRETQEIINKQVAYREWQSKLKEFQENTIKQLKELKMEIE
AIKKEHMETTLDIENOKKROGAVDTG"
BASE COUNT 390 a 300 c 272 g 151 t
ORIGIN
Query Match 2.6%; Score 28; DB 88; Length 1113;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 gctgtgtgtcaggagtgctgttagacct 83
|||||
Db 668 GCCTGTGTCTCAGGAGTGCTGTAGACCT 641
|||||

RESULT 5
AF000202/c
LOCUS AF000202 1117 bp DNA ROD 11-DEC-1997
DEFINITION Rattus sp. T-612 retrotransposon mlv12-rm6 5'UTR and putative RNA
binding protein gene, partial cds.
ACCESSION AF000202
VERSION AF000202.1 GI:2599312
KEYWORDS
SOURCE Rattus sp. T-612.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1117)
AUTHORS Cabot,E.L., Angeletti,B., Usdin,K. and Furano,A.V.
TITLE Rapid evolution of a young LI (LINE-1) clade in recently speciated
Rattus taxa
JOURNAL J. Mol. Evol. 45 (4), 412-423 (1997)
MEDLINE 97472947
REFERENCE 2 (bases 1 to 1117)
AUTHORS Cabot,E.L. and Furano,A.V.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA
FEATURES
source
location/Qualifiers
1..1117
/organism="Rattus sp. T-612"
/isolate="T-612"
/db_xref="taxon:68896"
/note="from the collection of Institut des Sciences de
l'Evolution, Montpellier II; collected from the island of
Ceram, Indonesia and identified by Djoko Iskandar as
Rattus moluccarius; cranial morphology suggests similarity
to Rattus rattus satarae"
repeat_region 1..>1117
/note="mlv12-rm6"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type=dispersed
1..733
734..>1117
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein"
/protein_id="AAC53497.1"
/db_xref="GI:2599313"

/note="L1mlvrn38b"
5'UTR
repeat_region
<1..729
1..>1113
/note="mlv12-rn38"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type=dispersed
730..>1113
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein 1"
/protein_id="AAB53049.1"
/db_xref="GI:2047337"
/translation="MARGKRNPNSNRNODCMPSSEPNSTKTNNMYPNTPEKODLVSK
SYLIMLEDFKDMNTLRETQEIINKQVAYREWQSKLKEFQENTIKQLKELKMEIE
AIKKEHMETTLDIENOKKROGAVDTG"
BASE COUNT 390 a 300 c 272 g 151 t
ORIGIN
Query Match 2.6%; Score 28; DB 88; Length 1113;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 gctgtgtgtcaggagtgctgttagacct 83
|||||
Db 668 GCCTGTGTCTCAGGAGTGCTGTAGACCT 641
|||||

RESULT 6
AF000200/c
LOCUS AF000200 1127 bp DNA ROD 11-DEC-1997
DEFINITION Rattus sp. T-612 retrotransposon mlv12-rm3 5'UTR and putative RNA
binding protein gene, partial cds.
ACCESSION AF000200
VERSION AF000200.1 GI:2599308
KEYWORDS
SOURCE Rattus sp. T-612.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1127)
AUTHORS Cabot,E.L., Angeletti,B., Usdin,K. and Furano,A.V.
TITLE Rapid evolution of a young LI (LINE-1) clade in recently speciated
Rattus taxa
JOURNAL J. Mol. Evol. 45 (4), 412-423 (1997)
MEDLINE 97472947
REFERENCE 2 (bases 1 to 1127)
AUTHORS Cabot,E.L. and Furano,A.V.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA
FEATURES
source
location/Qualifiers
1..1127
/organism="Rattus sp. T-612"
/isolate="T-612"
/db_xref="taxon:68896"
/note="from the collection of Institut des Sciences de
l'Evolution, Montpellier II; collected from the island of
Cerem, Indonesia and identified by Djoko Iskandar as
Rattus moluccarius; cranial morphology suggests similarity
to Rattus rattus satarae"
repeat_region 1..>1127
/note="mlv12-rm3"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type=dispersed
1..743
744..>1127
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein"
/protein_id="AAC53495.1"
/db_xref="GI:2599309"
/translation="MARGKRNPNSNRNODCMPSSEPNSTKTNNMYPNTPEKODLVSK
SYLIMLEDFKDMNTLRETQEIINKQVAYREWQSKLKEFQENTIKQLKELKMEIE
AIKKEHMETTLDIENOKKROGAVDTG"
BASE COUNT 390 a 308 c 273 g 156 t
ORIGIN
Query Match 2.6%; Score 28; DB 88; Length 1127;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 gctgtgtgtcaggagtgctgttagacct 83
```

```

|||||
Db 682 GCCTGTGTGCAGAGTCTGTAGACCT 655

RESULT 7
LOCUS AF000203 1127 bp DNA ROD 11-DEC-1997
DEFINITION Rattus sp. T-612 retrotransposon mlv12-rm8 5'UTR and putative RNA
binding protein gene, partial cds.
ACCESSION AF000203
VERSION AF000203.1 GI:2599314
KEYWORDS
SOURCE Rattus sp. T-612.
ORGANISM Rattus sp. T-612.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1127)
AUTHORS Cabot,E.L., Angeletti,B., Usdin,K. and Furano,A.V.
TITLE Rapid evolution of a young L1 (LINE-1) clade in recently speciated
Rattus taxa
J. Mol. Evol. 45 (4), 412-423 (1997)
MEDLINE 97472947
REFERENCE 2 (bases 1 to 1127)
AUTHORS Cabot,E.L. and Furano,A.V.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA
FEATURES
source
1. .1127
/organism="Rattus sp. T-612"
/isolate="T-612"
/db_xref="taxon:68896"
/note="from the collection of Institut des Sciences de
l'Evolution, Montpellier II; collected from the island of
Ceram, Indonesia and identified by Djoko Iskandar as
Rattus moluccarius; cranial morphology suggests similarity
to Rattus rattus satarae"
repeat_region
1. .>1127
/note="mlv12-rm8"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type="dispersed"
5'UTR
1. .743
CDS
744..>1127
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein"
/protein_id="AAC53498.1"
/db_xref="GI:2599315"
/translation="MARGKRRNPNNODCMPPSPNPTKTINMEYPTPEKQDLVSK
SYLLMLEDFFKDMNLTRETOELINKQVEAYREWQKSLAEFQENTIKQMKELKWEI
AIKKHMETLTDIENKRRQGAVDTS"
BASE COUNT 395 a 303 c 271 g 157 t 1 others
ORIGIN

Query Match 2.6%; Score 28; DB 88; Length 1127;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 gcctgtgtgcaggagtcgttagacct 83
|||||
Db 682 GCCTGTGTGCAGAGTCTGTAGACCT 655

RESULT 8
LOCUS AF000204 1297 bp DNA ROD 11-DEC-1997
DEFINITION Rattus sp. T-612 retrotransposon mlv12-rm13 5'UTR and putative RNA
binding protein gene, partial cds.
ACCESSION AF000204
VERSION AF000204.1 GI:2599316
KEYWORDS
SOURCE Rattus sp. T-612.
ORGANISM Rattus sp. T-612.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1297)
AUTHORS Cabot,E.L., Angeletti,B., Usdin,K. and Furano,A.V.
TITLE Rapid evolution of a young L1 (LINE-1) clade in recently speciated
Rattus taxa
J. Mol. Evol. 45 (4), 412-423 (1997)
MEDLINE 97472947
REFERENCE 2 (bases 1 to 1297)
AUTHORS Cabot,E.L. and Furano,A.V.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA
FEATURES
source
1. .1297
/organism="Rattus sp. T-612"
/isolate="T-612"
/db_xref="taxon:68896"
/note="from the collection of Institut des Sciences de
l'Evolution, Montpellier II; collected from the island of
Ceram, Indonesia and identified by Djoko Iskandar as
Rattus moluccarius; cranial morphology suggests similarity
to Rattus rattus satarae"
repeat_region
1. .>1297
/note="mlv12-rm13"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type="dispersed"
5'UTR
1. .913
CDS
914..>1297
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein"
/protein_id="AAC53499.1"
/db_xref="GI:2599317"
/translation="MARGKRRNPNNODCMPPSPNPTKTINMEYPTPEKQDLVSK
SYLLMLEDFFKDMNLTRETOELINKQVEAYREWQKSLAEFQENTIKQMKELKWEI
AIKKHMETLTDIENKRRQGAVDTS"
BASE COUNT 434 a 356 c 324 g 183 t
ORIGIN

Query Match 2.6%; Score 28; DB 88; Length 1297;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 gcctgtgtgcaggagtcgttagacct 83
|||||
Db 852 GCCTGTGTGCAGAGTCTGTAGACCT 825

RESULT 9
LOCUS RNL1MLV11/c 1297 bp DNA ROD 26-APR-1997
DEFINITION Rattus norvegicus L1 retrotransposon mlv12-rn11, 5'UTR and putative
RNA binding protein 1 gene, partial cds.
ACCESSION U87600
VERSION U87600.1 GI:2047330
KEYWORDS
SEGMENT 1 of 2
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1297)
AUTHORS Hayward,B.E., Zavanelli,M. and Furano,A.V.
TITLE Recombination creates novel L1 (LINE 1) elements in Rattus
norvegicus
Genetics (1997) In press
JOURNAL
REFERENCE 2 (bases 1 to 1297)

```

AUTHORS Hayward, B.E., Good, B., Cabot, E. and Furano, A.V.  
TITLE Direct Submission  
JOURNAL Submitted (29-JAN-1997) LMCB/NIDDK, NIH, 8/203, 8 Center Dr. MSC  
0830, Bethesda, MD 20892-0830, USA

FEATURES  
source  
1. .1297  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
5'UTR  
repeat\_region 1. .>1297  
/note="mlv12-rn11"  
/rpt\_family="LINE 1 non-LTR retrotransposable element"  
CDS  
914. .>1297  
/note="ORF 1"  
/codon\_start=1  
/product="putative RNA binding protein 1"  
/protein\_id="AAB53048.1"  
/db\_xref="GI:2047333"  
/translation="MARGKRRNPNRQDYMASSEPNSTKTNMEYPTPEKQDLVSK  
SYLIMLEDFKDLNLTREAOENINQKAYREESQSLKEFQENTIKRLKELMEIE  
AIKHEMETTLDIENOKRQGVDT"  
BASE COUNT 430 a 357 c 323 g 187 t  
ORIGIN  
|||||  
Query Match 2.6%; Score 28; DB 88; Length 1297;  
Best Local Similarity 100.0%; Pred. No. 0.0007;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 56 gctgtgtgtcaggagtgcttagacct 83  
|||||  
Db 852 GCCTGTGTCTCAGGAGTGCTGTAGACCT 825

RESULT 10  
AF000199/c  
LOCUS AF000199 1306 bp DNA 11-DEC-1997  
DEFINITION Rattus sp. T-612 retrotransposon u-rm2 5'UTR and putative RNA  
binding protein pseudogene, partial sequence.  
ACCESSION AF000199  
VERSION AF000199.1 GI:2599307  
KEYWORDS Rattus sp. T-612.  
SOURCE Rattus sp. T-612  
ORGANISM Rattus sp. T-612  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 1306)  
Cabot, E.L., Angeletti, B., Usdin, K. and Furano, A.V.  
Rapid evolution of a young L1 (LINE-1) clade in recently speciated  
Rattus taxa  
J. Mol. Evol. 45 (4), 412-423 (1997)  
MEDLINE 97472947  
REFERENCE 2 (bases 1 to 1306)  
AUTHORS Cabot, E.L. and Furano, A.V.  
TITLE Direct Submission  
JOURNAL Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8  
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA

FEATURES  
source  
1. .1306  
/organism="Rattus sp. T-612"  
/isolate="T-612"  
/db\_xref="taxon:68896"  
/note="from the collection of Institut des Sciences de  
Ceram, Indonesia and identified by Djoko Iskandar as  
Rattus moluccarius; cranial morphology suggests similarity  
to Rattus rattus sataarae"  
5'UTR  
repeat\_region 1. .>1306  
/note="u-rm2"  
/rpt\_family="LINE 1 non-LTR retrotransposable element"

CDS  
922. .>1306  
/note="ORF 1; putative RNA binding protein"  
/codon\_start=1  
/pseudo  
BASE COUNT 441 a 351 c 311 g 199 t 4 others  
ORIGIN  
|||||  
Query Match 2.6%; Score 28; DB 88; Length 1306;  
Best Local Similarity 100.0%; Pred. No. 0.00069;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 56 gctgtgtgtcaggagtgcttagacct 83  
|||||  
Db 860 GCCTGTGTCTCAGGAGTGCTGTAGACCT 833

RESULT 11  
RNM8LV11/c  
LOCUS RNM8LV11 1429 bp DNA 25-APR-1997  
DEFINITION Rattus norvegicus L1 retrotransposon mlv12-rn8, 5'UTR and putative  
RNA binding protein 1 gene, partial cds.  
ACCESSION U87598  
VERSION U87598.1 GI:2039337  
KEYWORDS 1 of 2  
SEGMENT Norway rat.  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 1429)  
Hayward, B.E., Zavanelli, M. and Furano, A.V.  
Recombination creates novel L1 (LINE 1) elements in Rattus  
norvegicus  
Genetics (1997) In press  
REFERENCE 2 (bases 1 to 1429)  
AUTHORS Hayward, B.E., Good, B., Cabot, E. and Furano, A.V.  
TITLE Direct Submission  
JOURNAL Submitted (29-JAN-1997) LMCB/NIDDK, NIH, 8/203 8 Center Dr. MSC  
0830, Bethesda, MD 20892-0830, USA

FEATURES  
source  
1. .1429  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
repeat\_region 1. .>1429  
/note="mlv12-rn8"  
/rpt\_family="LINE 1 non-LTR retrotransposable element"  
/rpt\_type="dispersed  
1. .910  
911. .>1429  
/note="ORF 1"  
/codon\_start=1  
/product="putative RNA binding protein 1"  
/protein\_id="AAB52995.1"  
/db\_xref="GI:2039340"  
/translation="MARGKRRNPNRQDYMASSEPNSTKTNMEYPTPEKQDLVSK  
SYLIMLEDFKDLNLTREAOENINQKAYREESQSLKEFQENTIKRLKELMEIEAIKHEMETTLDI  
ENOKRQGVDT"  
BASE COUNT 505 a 367 c 349 g 208 t  
ORIGIN  
|||||  
Query Match 2.6%; Score 28; DB 88; Length 1429;  
Best Local Similarity 100.0%; Pred. No. 0.00069;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 56 gctgtgtgtcaggagtgcttagacct 83  
|||||  
Db 849 GCCTGTGTCTCAGGAGTGCTGTAGACCT 822

```

RESULT 12
RNL1RTP5F/c
LOCUS RNL1RTP5F 1826 bp DNA ROD 29-JAN-1997
DEFINITION R.norvegicus L1 retroposon/pseudogene 5' flank and ORF1 (partial).
ACCESSION X61297
VERSION X61297.1 GI:56527
KEYWORDS 5' flanking region; L1 retroposon; pseudogene; reverse transcriptase-like protein.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1826)
AUTHORS Speek,M.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1991) M. Speek, Tartu University, Dept of Mol
Biol, 2 Jakobi St, Tartu 202 400, ESTONIA, USSR
REFERENCE 2 (bases 1 to 1826)
AUTHORS Ilves,H., Kahre,O. and Speek,M.
TITLE Translation of the rat LINE bicistronic RNAs in vitro involves ribosomal reinitiation instead of frameshifting
JOURNAL Mol. Cell. Biol. 12 (9), 4243-4248 (1992)
MEDLINE 92375092
COMMENT L1 protein homologous to the retroviral reverse transcriptase.
FEATURES
source
1..1826
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/dev_stage="adult"
/tissue_type="liver"
/clone_lib="genomic DNA in EMBL3"
/clone="GRM5"
misc_feature
1..727
/feature="5' flanking sequence"
repeat_region
713..727
/feature="insertion site for the L1"
promoter
728..1400
/rpt_type=DIRECT
/standard_name="L1"
CDS
1414..1824
/partial
/feature="L1 retroposon/pseudogene; beginning of ORF1
sequence"
/codon_start=1
/protein_id="CAA43596.1"
/db_xref="GI:582833"
/translation="PLSEIAQSNTRNLNLMARGKRRNPSNRNQCMPSPSPNPTKIN
MEYPTPEKQDLVSKSYLIMLDFKDKVNSLRQVEAYREESQSKLKEFQENTIKQ
LKEKMEIEAKKHEMETTLDIENKRRQAVDTs"
BASE COUNT 614 a 399 c 408 g 405 t
ORIGIN
Query Match 2.6%; Score 28; DB 88; Length 1826;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 gccgtgtgtcaggagtcgttagacct 83
|||||
Db 1397 GCCTGTGTGCAGGAGTGTGTAGACCT 1370
|||||

RESULT 13
RNLB6/c
LOCUS RNLB6 1857 bp DNA ROD 06-JUL-1989
DEFINITION Rat L1Rn B6 repetitive DNA element.
ACCESSION X07686
VERSION X07686.1 GI:56554
KEYWORDS L1 repetitive sequence; long interspersed repeat; repetitive sequence; tandem repeat.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1857)
AUTHORS Furano,A.V., Robb,S.M. and Robb,F.T.
TITLE The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements
JOURNAL Nucleic Acids Res. 16 (19), 9215-9231 (1988)
MEDLINE 89016625
FEATURES
source
1..1857
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/misc_feature
1..461
/feature="non L1 flanking DNA"
misc_feature
462..559
/feature="partial tandem duplication of L1Rn B6 DNA"
promoter
560..1167
/feature="L1Rn B6 DNA promoter"
CDS
1358..>1857
/feature="ORF1 (1857 is 2nd base in codon)"
/codon_start=1
/protein_id="CAA30528.1"
/db_xref="GI:56555"
/translation="MARGKRRNPSNRNQCMPSPSPNPTKINMEYPTPEKQDLVSK
SYLIMLDFKDKVNSLRQVEAYREESQSKLKEFQENTIKQVEAHREESQSKLKEF
QAVDTs"
repeat_region
1548..1679
/feature="direct repeat"
repeat_region
1680..1745
/feature="direct repeat"
BASE COUNT 623 a 462 c 418 g 354 t
ORIGIN
Query Match 2.6%; Score 28; DB 88; Length 1857;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 56 gccgtgtgtcaggagtcgttagacct 83
|||||
Db 1296 GCCTGTGTGCAGGAGTGTGTAGACCT 1269
|||||

RESULT 14
RNL4MLV1/c
LOCUS RNL4MLV1 1986 bp DNA ROD 25-APR-1997
DEFINITION Rattus norvegicus L1 retrotransposon mlv12-rn14, 5'UTR and putative RNA binding protein 1 gene, partial cds.
ACCESSION U87602
VERSION U87602.1 GI:2039341
KEYWORDS 1 of 2
SEGMENT
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1986)
AUTHORS Hayward,B.E., Zavanelli,M. and Furano,A.V.
TITLE Recombination creates novel L1 (LINE 1) elements in Rattus norvegicus
JOURNAL Genetics (1997) In press
REFERENCE 2 (bases 1 to 1986)
AUTHORS Hayward,B.E., Good,B., Cabot,E. and Furano,A.V.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1997) LMCB/NIDDK, NIH, 8/203, 8 Center Dr. MSC

```

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 1857)  
AUTHORS Furano,A.V.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1988) Furand A.V., NIH, 81203, Bethesda, MD 20892, USA

REFERENCE 2 (bases 1 to 1857)  
AUTHORS Furano,A.V., Robb,S.M. and Robb,F.T.  
TITLE The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements  
JOURNAL Nucleic Acids Res. 16 (19), 9215-9231 (1988)  
MEDLINE 89016625

FEATURES  
source  
1..1857 Location/Qualifiers  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
misc\_feature  
1..461  
/note="non L1 flanking DNA"  
misc\_feature  
462..559  
/note="partial tandem duplication of L1Rn B6 DNA"  
promoter  
560..1167  
/note="L1Rn B6 DNA promoter"  
CDS  
1358..>1857  
/note="ORF1 (1857 is 2nd base in codon)"  
/codon\_start=1  
/protein\_id="CAA30528.1"  
/db\_xref="GI:56555"

/translation="MARGKRRNPSNRNQCMPSPSPNPTKINMEYPTPEKQDLVSK  
SYLIMLDFKDKVNSLRQVEAYREESQSKLKEFQENTIKQVEAHREESQSKLKEF  
QAVDTs"  
repeat\_region  
1548..1679  
/note="direct repeat"  
repeat\_region  
1680..1745  
/note="direct repeat"  
BASE COUNT 623 a 462 c 418 g 354 t  
ORIGIN

Query Match 2.6%; Score 28; DB 88; Length 1857;  
Best Local Similarity 100.0%; Pred. No. 0.00068;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 gccgtgtgtcaggagtcgttagacct 83  
|||||  
Db 1296 GCCTGTGTGCAGGAGTGTGTAGACCT 1269  
|||||

RESULT 14  
RNL4MLV1/c  
LOCUS RNL4MLV1 1986 bp DNA ROD 25-APR-1997  
DEFINITION Rattus norvegicus L1 retrotransposon mlv12-rn14, 5'UTR and putative RNA binding protein 1 gene, partial cds.  
ACCESSION U87602  
VERSION U87602.1 GI:2039341  
KEYWORDS 1 of 2  
SEGMENT  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 1986)  
AUTHORS Hayward,B.E., Zavanelli,M. and Furano,A.V.  
TITLE Recombination creates novel L1 (LINE 1) elements in Rattus norvegicus  
JOURNAL Genetics (1997) In press  
REFERENCE 2 (bases 1 to 1986)  
AUTHORS Hayward,B.E., Good,B., Cabot,E. and Furano,A.V.  
TITLE Direct Submission  
JOURNAL Submitted (29-JAN-1997) LMCB/NIDDK, NIH, 8/203, 8 Center Dr. MSC

0830, Bethesda, MD 20892-0830, USA  
 1..1986  
 /location/Qualifiers  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 1..922  
 /note="mlv12-rn14"  
 /rpt\_family="LINE 1 non-LTR retrotransposable element"  
 /rpt\_type="dispersed"  
 923..>1986  
 repeat\_region 1..>1986  
 /codon\_start=1  
 /product="putative RNA binding protein 1"  
 /protein\_id="AAB52996.1"  
 /db\_xref="GI:2039344"  
 /translation="MARGKRRNPSSNRNDYMPSPSPSTKINMEYNTPEKQDLVSK  
 SLILMLDEDFKDMNTLREANINKQKAYREESQKSLKEFOENTIKQLKMEIE  
 AIKKHEMTTLDIENKRGAVDTFTNRQEMKXERISGAEDSIEIDSTVKDNVNR  
 XLLVQXIQEIXDSRRNRLIIGIESEDYQLGPNVFNKIIIEENFNLKYEIPID  
 IOEAVRTNRLDOKRNTSRHIIIVTPNAQNKERILKAVREKGVYTKGRPIRTPDFS  
 PETMKARSDVDVIQTLREHKQCPRLLYPAKLINIDGETKIFHDKTKFTQYLSTNPA  
 LQRTNGKAQHEASYILEEA"  
 BASE COUNT 733 a 499 c 439 g 309 t 6 others  
 ORIGIN

Query Match 2.6%; Score 28; DB 88; Length 1986;  
 Best Local Similarity 100.0%; Pred. No. 0.00068;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 56 gctgtgtgtcaggagtgctgttagacct 83  
 |||||  
 DB 861 GCCTGTGTGTGTCAGAGTGCTGTAGACCT 834

RESULT 15  
 AF058787  
 LOCUS AF058787 2225 bp mRNA ROD 19-APR-1998  
 DEFINITION Rattus norvegicus heme oxygenase-3 (HO-3) mRNA, complete cds.  
 ACCESSION AF058787  
 VERSION AF058787.1 GI:3063688  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 2225)  
 McCoubrey, W.K. Jr., Huang, T.J. and Maines, M.D.  
 Isolation and characterization of a cDNA from the rat brain that  
 encodes hemoprotein heme oxygenase-3  
 Eur. J. Biochem. 247 (2): 725-732 (1997)  
 97410004  
 MEDLINE  
 REFERENCE 2 (bases 1 to 2225)  
 McCoubrey, W.K. Jr., Huang, T.J. and Maines, M.D.  
 Direct Submission  
 AUTHORS  
 TITLE  
 JOURNAL  
 Submitted (09-APR-1998) Biochemistry and Biophysics, University of  
 Rochester, 601 Elmwood Avenue Box 712, Rochester, NY 14642, USA  
 Location/Qualifiers  
 1..2225  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /tissue\_type="brain"  
 1..2225  
 /gene="HO-3"  
 1062..1934  
 /gene="HO-3"  
 /codon\_start=1  
 /product="heme oxygenase-3"  
 /protein\_id="AAC1442.1"  
 /db\_xref="GI:3063688"  
 /translation="MSSEVETAEVDESEKNSMAKSHKSIADFSLLKEGTKEAD  
 DRAENTQFVKDFLKGNIKELFKLATTALSAPPEEMDSLTKDMEYFFGENWEEKVK

CSEAAQYYVDQIHVQGNPEHLVAHTYTYMGNNLSGQVLKKETOPVPFTREGTQF  
 YLFHYDNAQKFLFYCARLNALDLNLKTKRIVEEATKAFVNMQIRSELDQAGSIP  
 VRETLKNGLSILDGKGVCKCPFNAAQPKDKTILGSGNCFQMSMALLRPNQLILYA  
 :SMALVAGLLAWYTM"  
 BASE COUNT 539 a 509 c 552 g 625 t  
 ORIGIN

Query Match 2.6%; Score 28; DB 88; Length 2225;  
 Best Local Similarity 100.0%; Pred. No. 0.00067;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 56 gctgtgtgtcaggagtgctgttagacct 83  
 |||||  
 DB 296 GCCTGTGTGTGTCAGAGTGCTGTAGACCT 323

Search completed: November 4, 2000, 04:20:11  
 Job time: 22321 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:39:12 ; Search time 255.06 Seconds  
(without alignments)  
1605.396 Million cell updates/sec

Title: US-09-101-423A-6  
Perfect score: 1090  
Sequence: 1 ttgtctgtgttacttgt.....cctagcagctgcatccaca 1090

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 480022 seqs, 187831343 residues

Word size : 0

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_36:\*

- 1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT:\*
- 2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT:\*
- 3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT:\*
- 4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT:\*
- 5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT:\*
- 6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT:\*
- 7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT:\*
- 8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT:\*
- 9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT:\*
- 10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT:\*
- 11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT:\*
- 12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT:\*
- 13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT:\*
- 14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT:\*
- 15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT:\*
- 16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT:\*
- 17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT:\*
- 18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:\*
- 19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:\*
- 20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:\*
- 21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1090	100.0	1090	18 T72790	Metastasis inducin
2	28	2.6	437	15 Q55636	P35 sequence of hu
3	28	2.6	437	15 Q55637	Rat retrotranspos
4	28	2.6	2225	21 294380	Rat haem oxygenase
5	28	2.6	44576	21 Z61522	Cosmid CVO14 conta
6	19	1.7	1374	15 Q72874	Phosphatase DbpA.
7	19	1.7	1614	15 Q72875	Phosphatase DbpA.
8	18	1.7	356	20 X97955	Human secreted pro
9	18	1.7	1024	21 A01797	Human colon cancer
10	18	1.7	1558	20 X90849	CDNA clone ct547_2
11	18	1.7	32207	20 V73805	KSHV LUR DNA (nucl
12	18	1.7	51259	18 X83007	Partial mouse WRN

13	18	1.7	137507	19 V19941	KSHV long unique c
14	17	1.6	831	16 Q79912	Aromatic dihydrodi
15	17	1.6	1569	5 N40050	Sequence of 'a'-fa
16	17	1.6	2244	17 T10429	Mature isoamylase
17	17	1.6	2244	19 V23640	Flavobacterium iso
18	17	1.6	2334	17 T10428	Isoamylase gene.
19	17	1.6	2334	19 V23639	Full length Flavob
20	17	1.6	2473	19 X14055	H. pylori GHP0 139
21	17	1.6	4531	20 X08688	Novel nucleotide s
22	17	1.6	6360	13 Q20577	Human Cytomegalovi
23	17	1.6	7294	20 X13000	Enterococcus faeca
24	17	1.6	12438	20 X13110	Enterococcus faeca
25	17	1.6	26281	21 Z60929	Nucleotide sequenc
26	17	1.6	37856	21 A11992	S. cellulosum DNA
27	17	1.6	138159	21 A34791	Human adenosine re
28	17	1.6	141589	21 A35005	Human adenosine re
29	17	1.6	141589	21 A35030	Human adenosine re
30	17	1.6	1664976	19 V21209	Methanococcus jann
31	16	1.5	300	20 Z14802	Human gene express
32	16	1.5	307	21 A31654	Plant microsatelli
33	16	1.5	327	18 T72464	Human CII-3 gene 5
34	16	1.5	350	16 T23970	Human gene signatu
35	16	1.5	352	20 Z42203	Human normal bladd
36	16	1.5	397	14 Q60786	Human brain expres
37	16	1.5	439	20 Z42200	Human normal bladd
38	16	1.5	440	20 X84178	Human normal bladd
39	16	1.5	477	21 Z45628	DNA encoding a mar
40	16	1.5	615	18 T90427	K-ras genomic frag
41	16	1.5	669	20 Z42166	Human normal bladd
42	16	1.5	669	21 Z53895	Neisseria meningit
43	16	1.5	669	21 Z54586	Neisseria meningit
44	16	1.5	708	21 Z45618	DNA encoding a mar
45	16	1.5	768	20 Z17223	Human gene express

#### ALIGNMENTS

RESULT 1

T72790

ID T72790 standard; DNA; 1090 BP.

XX

AC T72790;

XX

DT 22-SEP-1997 (first entry)

XX

DE Metastasis inducing DNA C20.

XX

KW Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;

KW osteopontin; ss.

XX

OS Homo sapiens.

XX

PN WO9725443-A1.

XX

PD 17-JUL-1997.

XX

PF 10-JAN-1997; 97WO-GB000074.

XX

PR 10-JAN-1996; 96GB-0000470.

XX

(UYLI-) UNIV LIVERPOOL.

PA Barracough BR, Rudland PS;

PI WPI; 1997-372878/34.

DR New isolated metastasis-inducing DNA - used to develop products to

XX identify and treat patients at risk from metastatic tumours

PT Claim 13; Page 27; 38pp; English.

PS Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12

CC

CC and C20 (72785-90) are entirely novel short stretches of human  
CC regulatory DNA capable of inducing metastasis. They were  
CC identified using a method for detecting Met-DNA that involves  
CC transferring human DNA from a malignant, metastatic cancer cell  
CC (in this case, breast cancer) into a cell line (pref. rat Rama 37)  
CC that produces only benign, non-metastasizing tumours when injected  
CC into a syngeneic animal, injecting those transformed cells into a  
CC syngeneic animal, selecting those animals having metastasizing  
CC tumours, and recovering the Met-DNA from them. The isolated  
CC Met-DNAs can be used to develop products to identify and treat  
CC patients at risk from metastatic tumours.

XX  
SQ Sequence 1090 BP; 224 A; 277 C; 275 G; 314 T; 0 other;

Query Match 100.0%; Score 1090; DB 18; Length 1090;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1090; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtctctggtgttactgttttccattctgacagtggtttgaccttctatagcctg 60  
DB 1 ttgtctctggtgttactgttttccattctgacagtggtttgaccttctatagcctg 60

QY 61 tgtgtcaggagtgctgttagacctattttctgtttttcttccagcagttacaggaaacaga 120  
DB 61 tgtgtcaggagtgctgttagacctattttctgtttttcttccagcagttacaggaaacaga 120

QY 121 ggtgttctactgtcagatgtgttagctgttctctcactgaactttcaagctgtctctgtgt 180  
DB 121 ggtgttctactgtcagatgtgttagctgttctctcactgaactttcaagctgtctctgtgt 180

QY 181 gcaggaaaccagagggcctctctacttctactctggtgcccctacacagggggcctaga 240  
DB 181 gcaggaaaccagagggcctctctacttctactctggtgcccctacacagggggcctaga 240

QY 241 tgggtctagggtgtttctctctagagcctgaatgtgggcagagagtagtctctctgtgt 300  
DB 241 tgggtctagggtgtttctctctagagcctgaatgtgggcagagagtagtctctctgtgt 300

QY 301 tcttaggtatgtctccctctgaaggtctagctctctctctccctccatgggatatgggtgag 360  
DB 301 tcttaggtatgtctccctctgaaggtctagctctctctctccctccatgggatatgggtgag 360

QY 361 ggagctgtttgaccaggtctctcctcctcctcctcctcctcctcctcctcctcctcctcct 420  
DB 361 ggagctgtttgaccaggtctctcctcctcctcctcctcctcctcctcctcctcctcctcct 420

QY 421 cttgctctgctgcaatctctccacacacacagagggcaccacagttctctctgggccaagat 480  
DB 421 cttgctctgctgcaatctctccacacacacagagggcaccacagttctctctgggccaagat 480

QY 481 gtgggcaaaaggtggcagaagtggaatctctcctcctcctcctcctcctcctcctcctcctc 540  
DB 481 gtgggcaaaaggtggcagaagtggaatctctcctcctcctcctcctcctcctcctcctcctc 540

QY 541 actcttgggcaatccgctctctctcctcctcctcctcctcctcctcctcctcctcctcctcct 600  
DB 541 actcttgggcaatccgctctctctcctcctcctcctcctcctcctcctcctcctcctcctcct 600

QY 601 atcaggcaaaaggtttgaggaacacagttagaactggaggtcaggtccacagagaatt 660  
DB 601 atcaggcaaaaggtttgaggaacacagttagaactggaggtcaggtccacagagaatt 660

QY 661 ttgctcttctgtgtcctgagtcacacagcaggtcacttgagcagaaaaattggttttc 720  
DB 661 ttgctcttctgtgtcctgagtcacacagcaggtcacttgagcagaaaaattggttttc 720

QY 721 cctcgtgtctcaggtgaaggtgacactcaggtgtgtgtttcagctgtacctgtggaaa 780  
DB 721 cctcgtgtctcaggtgaaggtgacactcaggtgtgtgtttcagctgtacctgtggaaa 780

QY 781 gtatgtgttttaaaatcaagatagctatcatgacagaggtctgtgttaaatgtctatt 840  
DB 781 gtatgtgttttaaaatcaagatagctatcatgacagaggtctgtgttaaatgtctatt 840

DB 781 gtatgtgttttaaaatcaagatagctatcatgacagaggtctgtgttaaatgtctatt 840  
QY 841 tgggtctctttagacttacttcttctgttactgagagatcaaaactagggtctcacaagctc 900  
DB 841 tgggtctctttagacttacttcttctgttactgagagatcaaaactagggtctcacaagctc 900  
QY 901 atcacaaatcctgtcactgacgtccagctccattcttcttcttcttcttcttcttcttcttct 960  
DB 901 atcacaaatcctgtcactgacgtccagctccattcttcttcttcttcttcttcttcttcttct 960  
QY 961 ctgcgcagcaagaaacacgcgttaggacatagcagatccttctgctgcagccaaactttat 1020  
DB 961 ctgcgcagcaagaaacacgcgttaggacatagcagatccttctgctgcagccaaactttat 1020  
QY 1021 tgaatcttaaggagagccgcgcagcagctggtggtggtgttttatatacacccctagcacag 1080  
DB 1021 tgaatcttaaggagagccgcgcagcagctggtggtggtgttttatatacacccctagcacag 1080  
QY 1081 tgcattccaca 1090  
DB 1081 tgcattccaca 1090

RESULT 2  
Q65636/c  
ID Q65636 standard; DNA; 437 BP.

XX Q65636;  
XX  
XX 18-JAN-1995 (first entry)  
XX  
XX P35 sequence of human prostate cancer.  
XX  
XX Testing; detection; identification; polymorphism; analysis; cancer;  
XX pathogen; pedigree; inherited disorders; mutation; ss.  
XX  
XX Homo sapiens.  
XX  
XX W09411383-A.  
XX  
XX 26-MAY-1994.  
XX  
XX 08-NOV-1993; 33WO-US10722.  
XX  
XX 12-NOV-1992; 92US-0974447.  
XX  
XX (COLD-) COLD SPRING HARBOR LAB.  
XX  
XX Lisitsyn N, Wigler M;  
XX  
XX WPI; 1994-183414/22.  
XX  
XX Prodn. of probes for detecting DNA sequence differences - using  
XX DNA from 2 related different sources and representational  
XX difference analysis  
XX  
XX Disclosure; Figure 3; 47pp; English.  
XX  
XX The P35 human prostate cancer sequence was isolated using an  
XX identification/amplification method known as representational  
XX difference analysis.  
XX  
XX Sequence 437 BP; 156 A; 121 C; 97 G; 63 T; 0 other;

Query Match 2.6%; Score 28; DB 15; Length 437;  
Best Local Similarity 100.0%; Pred. No. 0.00024;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 gctctgtgtgt-agaggtctctagacct 83  
DB 255 GCCGTGTGTAGGAGTCTGTAGACCT 228

us-09-101-423a-6.olig.rng

Mon Nov 6 10:14:41 2000

```

XX PD 09-MAR-2000.
XX PF 27-AUG-1999; 99WO-US19823.
XX XX 28-AUG-1998; 98US-0098377.
PR PR 25-FEB-1999; 99US-0121946.
XX XX (HARD ) HARVARD COLLEGE.
XX PA (LEEM/) LEE M.
XX PA (PERR/) PERRELLA M A.
XX PA (YETS/) YET S.
XX PI Lee M, Perrella MA, Yet S;
XX XX WPI; 2000-256492/22.
XX DR P-PSDB; Y79283.
XX XX Inhibiting cardiomyocyte death by administering a haem oxygenase is
XX PT useful to treat cardiovascular injury and to preserve a heart for
XX PT transplant purposes
XX XX Disclosure; Page 19-21; 27pp; English.
XX XX The present sequence is that of cDNA encoding rat haem oxygenase
XX CC HO-3 (see Y79285). The invention features methods of minimizing
XX CC myocardial damage by salvaging hypoxic myocardial tissue before it
XX CC becomes irreversibly injured. A method of inhibiting
XX CC cardiomyocyte death in a mammal, e.g. a human, who has suffered a
XX CC myocardial infarction or who has myocarditis, is carried out by
XX CC administering to the myocardium a HO polypeptide, preferably
XX CC selected from HO-1, HO-2 or HO-3, or by administering DNA encoding
XX CC an HO. Cardiomyocyte death is inhibited in vitro by contacting
XX CC cardiomyocytes with HO or DNA encoding HO. For example, isolated
XX CC myocardial tissue, such as a donor heart to be used in
XX CC transplantation, is preserved for up to 24 hr by bathing or
XX CC perfusing with a solution containing HO or DNA containing HO.
XX CC Vascular smooth muscle cell proliferation is inhibited by
XX CC administering HO to injured vascular tissue within 24 hr, or up to
XX CC 1 wk, after injury.
XX XX Sequence 2225 EP; 539 A; 509 C; 552 G; 625 T; 0 other;
XX SQ

Query Match 2.6%; Score 28; DB 21; Length 2225;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 gctgtgtcagagtgctgtacacct 83
DB 296 gctgtgtcagagtgctgtacacct 323

RESULT 5
Z61522/C
ID Z61522 standard; DNA; 44576 BP.
XX AC Z61522;
XX XX 19-JUN-2000 (first entry)
XX DT
XX DE Cosmid CVO14 containing rat vasopressin and oxytocin locus.
XX XX Oxytocin expressed sequence tag; 5'-OT EST; obesity; fertility; male;
XX KW transgenic animal; human late onset obesity; late onset visceral obesity;
XX KW male infertility; wasting; anorexia; cachexia; malabsorptive state;
XX KW catabolic state; inflammatory condition; Crohn's disease; AIDS wasting;
XX KW burn; cancer; bone disease; vasopressin; oxytocin; ss.
XX XX Rattus sp.
XX OS
XX PN WO200009686-A1..
XX XX

```

```

XX PD 09-MAR-2000.
XX PF 27-AUG-1999; 99WO-US19823.
XX XX 28-AUG-1998; 98US-0098377.
PR PR 25-FEB-1999; 99US-0121946.
XX XX (HARD ) HARVARD COLLEGE.
XX PA (LEEM/) LEE M.
XX PA (PERR/) PERRELLA M A.
XX PA (YETS/) YET S.
XX PI Lee M, Perrella MA, Yet S;
XX XX WPI; 2000-256492/22.
XX DR P-PSDB; Y79283.
XX XX Inhibiting cardiomyocyte death by administering a haem oxygenase is
XX PT useful to treat cardiovascular injury and to preserve a heart for
XX PT transplant purposes
XX XX Disclosure; Page 19-21; 27pp; English.
XX XX The present sequence is that of cDNA encoding rat haem oxygenase
XX CC HO-3 (see Y79285). The invention features methods of minimizing
XX CC myocardial damage by salvaging hypoxic myocardial tissue before it
XX CC becomes irreversibly injured. A method of inhibiting
XX CC cardiomyocyte death in a mammal, e.g. a human, who has suffered a
XX CC myocardial infarction or who has myocarditis, is carried out by
XX CC administering to the myocardium a HO polypeptide, preferably
XX CC selected from HO-1, HO-2 or HO-3, or by administering DNA encoding
XX CC an HO. Cardiomyocyte death is inhibited in vitro by contacting
XX CC cardiomyocytes with HO or DNA encoding HO. For example, isolated
XX CC myocardial tissue, such as a donor heart to be used in
XX CC transplantation, is preserved for up to 24 hr by bathing or
XX CC perfusing with a solution containing HO or DNA containing HO.
XX CC Vascular smooth muscle cell proliferation is inhibited by
XX CC administering HO to injured vascular tissue within 24 hr, or up to
XX CC 1 wk, after injury.
XX XX Sequence 2225 EP; 539 A; 509 C; 552 G; 625 T; 0 other;
XX SQ

Query Match 2.6%; Score 28; DB 15; Length 437;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 gctgtgtcagagtgctgtacacct 83
DB 255 GCCTGTGTCTCAGGAGTGCTGTGACCT 228

RESULT 4
Z94380
ID Z94380 standard; cDNA; 2225 BP.
XX AC Z94380;
XX XX 03-JUL-2000 (first entry)
XX DT
XX DE Rat haem oxygenase HO-3 cDNA.
XX XX HO-3; haem oxygenase; rat; heart transplant; preservation;
XX KW stenosis; restenosis; myocardial infarction; myocarditis;
XX KW vascular injury; gene therapy; antisense therapy; ss.
XX XX Rattus rattus.
XX XX Key Location/Qualifiers
XX FH 1062..1934
XX FT /*tag= a
XX PN WO200012118-A2.

```



us-09-101-423a-6.olig.rng

Mon Nov 6 10:14:41 2000

Carter KC, Duan RD, Feng P, Ferrie AM, Florence C;  
Florence K, Greene JM, Janat F, Kyaw H, Moore PA;  
N1 J, Rosen CA, Ruben SM, Shi Y, Soppet DR, Wei Y;  
Yu G;

WPI: 1999-418749/35.  
P-FSD8; Y36263

New isolated human genes encoding secreted polypeptides

Claim 1; Page 295-296; 537pp; English.

XX X97916 to X98029 represent 110 isolated human secreted protein genes.  
XX Y36224 to Y36727 represent the secreted proteins encoded by the 110  
CC human genes. The genes and their corresponding secreted polypeptides are  
CC useful for preventing, treating or ameliorating medical conditions,  
CC e.g. by protein or gene therapy. Also pathological conditions can be  
CC diagnosed by determining the amount of the new polypeptides in a sample  
CC or by determining the presence of mutations in the new genes. Specific  
CC uses are described for each of the 110 genes, based on which tissues they  
CC are most highly expressed in, and include developing products for the  
CC diagnosis or treatment of cancer, tumours, developmental abnormalities  
CC and foetal deficiencies, blood disorders, diseases of the immune system,  
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive  
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin  
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney  
CC disorders, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
CC The sequences given in X97907 to X97915 and Y36223 are used in the  
CC exemplification of the present invention.

XX Sequence 356 BF: 113 A; 84 C; 46 G; 113 T; 0 other;

Query Match 1.7%; Score 18; DB 20; Length 356;

Best Local Similarity 100.0%; Pred. No. 22; Mismatches 0; Indels 0; Gaps 0;

QY 17 ttgtttccattctga 34  
Db 113 ttgtttccattctga 130

RESULT 9  
A01797/c  
ID A01797 standard; CDNA; 1024 BP.

XX A01797;  
AC A01797;  
XX 19-MAY-2000 (first entry)

Human colon cancer cell line polynucleotide sequence SEQ ID NO:1788.

Human; colon cancer; tumour; diagnosis; gene expression product;  
probe; detection; cancerous state; metastasis; identification;  
breast cancer; oestrogen receptor-positive breast cancer; therapy;  
oestrogen receptor-negative breast cancer; lung cancer; ss.

XX Homo sapiens.

OS WO9958675-A2.

PN 18-NOV-1999.

PD 13-MAY-1999; 99WO-US10602.

PF 14-MAY-1998; 98US-0085426.

PR 15-MAY-1998; 98US-0085537.

PR 15-MAY-1998; 98US-0085696.

PR 21-OCT-1998; 98US-0105234.

PR 27-OCT-1998; 98US-0105877.

(CHIR ) CHIRON CORP.

PA (UYDU-) UNIV DUNDEE.

PI Fuller-Pace FV, Lane DP;

XX WPI: 1994-333217/41.

XX Detecting antimicrobial activity from effect on microbial  
PT nucleotide phosphatase - also new antimicrobial agents and  
PT recombinant DNA encoding the phosphatase D6pA, specific for  
PT prokaryotic ribosomal RNA.

XX Disclosure; Page 39-42; 57pp; English.

XX An RNA-helicase, DbpA, is used to screen for antimicrobial agents.  
CC Inhibition of DbpA would impede prokaryotic ribosomal function  
CC without affecting eukaryotic host cell RNAs. 2 DNA sequences  
CC encoding DbpA are given in Q72874-75. The second, containing  
CC additional flanking sequences, has been deposited as NCRC 12867  
CC as part of a plasmid contained within E. coli DH5-alpha.

XX Sequence 1614 BP; 379 A; 392 C; 451 G; 392 T; 0 other;

Query Match 1.7%; Score 19; DB 15; Length 1614;

Best Local Similarity 100.0%; Pred. No. 6.8; Mismatches 0; Indels 0; Gaps 0;

QY 651 cagaggaatttccttg 669  
Db 822 CAGAGGAATTTGCTTG 804

RESULT 8

XX X97955  
ID X97955 standard; DNA; 356 BP.

XX X97955;

XX 17-SEP-1999 (first entry)

XX Human secreted protein gene 40.

XX Human; secreted protein; cancer; tumour; developmental abnormality;  
KW foetal deficiency; blood disorder; immune system disorder; inflammation;  
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;  
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;  
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;  
KW digestive disorder; endocrine disorder; infection; AIDS; ss.

XX Homo sapiens.

OS WO9931117-A1.

PN 24-JUN-1999.

PD 17-DEC-1998; 98WO-US27059.

PF 19-DEC-1997; 97US-0068369.

PR 18-DEC-1997; 97US-0068006.

PR 18-DEC-1997; 97US-0068007.

PR 18-DEC-1997; 97US-0068008.

PR 18-DEC-1997; 97US-0068009.

PR 18-DEC-1997; 97US-0068053.

PR 18-DEC-1997; 97US-0068054.

PR 18-DEC-1997; 97US-0068057.

PR 18-DEC-1997; 97US-0068064.

PR 18-DEC-1997; 97US-0070923.

PR 19-DEC-1997; 97US-0068169.

PR 19-DEC-1997; 97US-0068365.

PR 19-DEC-1997; 97US-0068367.

PR 19-DEC-1997; 97US-0068368.

(HUMA-) HUMAN GENOME SCI INC.

(HYSE-) HYSEQ INC.

PA Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
 PI Lanson G, Drmanac R, Cravenjakov R, Dickson M, Drmanac S, Labat I;  
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
 XX WPI; 2000-126369/11.

Polynucleotide library used to determine cancerous states of mammalian cells -

Claim 1; Page 661; 1097pp; English.

CC A00010 to A02716 represent polynucleotides isolated from cDNA libraries  
 CC constructed from human colon cancer cell lines. The present invention  
 CC also describes a method of detecting differentially expressed genes  
 CC correlated with a cancerous state of a mammalian cell, comprising  
 CC detecting at least one differentially expressed gene product in a test  
 CC sample derived from a cell suspected of being cancerous, where detection  
 CC of the differentially expressed gene product is correlated with a  
 CC cancerous state of the cell from which the test sample was derived.  
 CC The polynucleotides sequences can be used in a method for detecting  
 CC differentially expressed genes correlated with a cancerous state of a  
 CC mammalian cell. The polynucleotides can also be used as probes for  
 CC detecting and mapping related genes. They can be used in diagnosis and  
 CC prognosis of diseases and disorders (e.g. identification of  
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or  
 CC responsiveness of cancer to therapy). This is particularly for breast  
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
 CC negative breast cancer, lung cancer, and colon cancer.

SQ Sequence 1024 BP; 274 A; 145 C; 181 G; 294 T; 130 other;

Query Match 1.7%; Score 18; DB 21; Length 1024;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 705 agaaaaattggtttccc 722  
 Db |||||  
 596 AGAAAAATTGGTTTCCC 579

RESULT 10  
 X90849  
 ID X90849 standard; DNA; 1558 BP.

AC X90849;

DT 17-JAN-2000 (first entry)

DE cDNA clone ct547\_2.

KW clone ct547\_2; ct547\_2 protein; human adult brain cDNA library;  
 KW secreted protein; AA428546; human cDNA clone 773189 5'; gene therapy;  
 KW cytokine; nutritional activity; cell proliferation; immune stimulation;  
 KW immune suppression; hematopoiesis regulation; tissue growth;  
 KW tumour inhibition; ds.

OS Homo sapiens.

EH Key Location/Qualifiers

FT CDS 202..417

FT /tag= a

FT /product= "ct547\_2 protein"

FT sig\_peptide 202..369

FT /tag= b

FT mat\_peptide 370..417

FT /tag= c

FT /note= "Mature ct547\_2 protein"

XX W09950405-A1.

XX PD

XX 07-OCT-1999.

XX 30-MAR-1999; 99WO-0505946.

XX 31-MAR-1998; 98US-0080110.

XX 29-MAR-1999; 99US-0280591.

XX (GEM) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX Merberg D, Treacy M, Agostino MJ, Steininger RJ;

XX WPI; 1999-610849/52.

XX P-PSDB; Y28807.

XX Polynucleotides encoding secreted human proteins, derived from human

XX adult brain, human fetal brain, human fetal kidney, and human adult

XX blood cDNA libraries

XX Claim 12; Page 99; 122pp; English.

XX The present nucleotide sequence comprises the full-length protein-coding

XX sequence of clone ct547\_2. ct547\_2 was isolated from a human adult brain

XX cDNA library using methods specific for secreted protein cDNAs. This

XX demonstrates some similarity with AA428546 of human cDNA clone 773189 5'.

XX This can be used in gene therapy. The polynucleotide and protein may

XX effect nutritional activity, cytokine and cell proliferation, immune

XX stimulation or suppression, hematopoiesis regulation, tissue growth,

XX tumour inhibition etc.

XX Sequence 1558 BP; 467 A; 287 C; 264 G; 540 T; 0 other;

Query Match 1.7%; Score 18; DB 20; Length 1558;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 ttgttttccattctga 34

Db |||||

668 ttgttttccattctga 685

RESULT 11

V73805

ID V73805 standard; DNA; 32207 BP.

AC V73805;

DT 25-FEB-1999 (first entry)

DE KSHV LUR DNA (nucleotides 105,301-137,507).

KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;

KW dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;

KW diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;

KW glycoprotein; Kaposin; cyclin D; immediate early protein; IEP; OX-2;

KW v-adh; G-protein coupled receptor; FGFRAT; ds.

OS Kaposi's sarcoma-associated herpesvirus.

XX US5849564-A.

XX 15-DEC-1998.

XX 29-NOV-1996; 96US-0770379.

XX 29-NOV-1996; 96US-0770379.

XX (UYCO ) UNIV COLOMBIA NEW YORK.

XX Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;

XX



DR WPI; 1999-069741/06.

XX Kaposi's sarcoma-associated herpes virus nucleic acid - encodes

PT dihydrofolate reductase and is useful for treatment, prophylaxis

PT or diagnosis of Kaposi's sarcoma

XX PS Disclosure; Column 155-182; 109pp; English.

XX This sequence is a fragment of the Kaposi's sarcoma-associated

CC herpesvirus (KSHV) LUR (long unique region). This fragment contains

CC coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67

CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,

CC ORF69, K12 which encodes Kaposin, K13, ORF72 which encodes cyclin D,

CC ORF73 which encodes immediate early protein (IEP), K14 which encodes

CC OX-2 (v-ach), ORF74 which encodes G-protein coupled receptor, ORF75

CC which encodes tegument protein/FCARAT, K15. KSHV is a new human

CC Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the

CC most common form of neoplasm occurring in persons with acquired immune

CC deficiency syndrome (AIDS). The dhfr protein is useful for vaccination,

CC prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma

CC and for detecting expression of a DNA virus associated with Kaposi's

CC sarcoma in a cell.

XX SQ Sequence 32207 BP; 7229 A; 9156 C; 8713 G; 7109 T; 0 other;

Query Match 1.7%; Score 18; DB 20; Length 32207;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 443 caccagagggcaccgaag 460

Db 10614 caccagagggcaccgaag 10631

RESULT 12

XX X83007 standard; DNA; 51259 BP.

XX X83007;

XX 31-AUG-1999 (first entry)

XX Partial mouse WRN genomic sequence #3.

XX Mouse; WRN; Werner's syndrome; detection; diagnosis; autosomal;

KW recessive disorder; phenotype; ss.

XX Mus musculus.

XX WO9724435-A1.

XX 10-JUL-1997.

XX 30-DEC-1996; 96WO-US20785.

XX 12-APR-1996; 96US-0632175.

XX 29-DEC-1995; 95US-0009409.

XX 29-DEC-1995; 95US-0580539.

XX 30-JAN-1996; 96US-0010835.

XX 30-JAN-1996; 96US-0594242.

XX (DARW-) DARWIN MOLECULAR CORP.

PA (OSHI/) OSHIMA J.

XX Fu Y, Mulligan J, Schellenberg GD, Yu C;

XX WPI; 1997-363671/33.

XX Isolated nucleic acid molecule encoding the WRN gene product

PT useful for detection and treatment of Werner's syndrome, and related

PT diseases

PS Claim 1; Fig 7; 153pp; English.

XX This sequence represents a fragment of the genomic sequence containing

CC the coding region for the mouse WRN gene (X83004). The corresponding

CC human gene (X83001) encodes a protein related to Werner's syndrome.

CC The products can be used for the detection and treatment of Werner's

CC syndrome (WS), an autosomal recessive disorder with a complex phenotype,

CC as well as related diseases.

XX SQ Sequence 51259 BP; 14533 A; 9635 C; 10266 G; 16825 T; 0 other;

Query Match 1.7%; Score 18; DB 18; Length 51259;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 129 ctgtcagatgtgtagctg 146

Db 50327 ctgtcagatgtgtagctg 50344

RESULT 13

XX V19941

XX V19941 standard; DNA; 137507 BP.

XX V19941;

XX 03-AUG-1998 (first entry)

XX KSHV long unique coding region and terminal repeat.

XX KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;

KW interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;

KW complement-binding protein; glycoprotein; capsid protein IV; infection;

KW immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;

KW lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;

KW HIV immune status; anti-inflammatory agent; therapy; ds.

XX OS Kaposi's sarcoma-associated herpes virus.

XX Key Location/Qualifiers

CDS 1142..2794

FT /\*tag= a

FT /product= complement-binding protein

CDS 8699..11236

FT /\*tag= b

FT /product= glycoprotein B

CDS complement (17261..17875)

FT /\*tag= c

FT /product= interleukin 6

CDS complement (21548..21832)

FT /\*tag= d

FT /product= macrophage inflammatory protein II

CDS complement (27137..27424)

FT /\*tag= e

FT /product= interferon regulatory factor 1

CDS 28661..29741

FT /\*tag= f

FT /product= protein T1.1

CDS complement (58976..60175)

FT /\*tag= g

FT /product= glycoprotein M

CDS complement (69412..69915)

FT /\*tag= h

FT /product= glycoprotein L

CDS complement (88410..88910)

FT /\*tag= i

FT /product= interferon regulatory factor 2

CDS 89600..90541

FT /\*tag= j

FT /product= interferon regulatory factor 3

CDS 90173..90643

FT /\*tag= k

FT CDS /product= glycoprotein x  
 complement (93636..94127)  
 /tag= 1  
 FT CDS /product= interferon regulatory factor 4  
 complement (111931..112443)  
 /tag= m  
 FT CDS /product= capsid protein IV  
 complement (123808..127296)  
 /tag= n  
 FT /product= immediate early protein

W09804576-A1.

05-FEB-1998.

22-JUL-1997; 97WO-US13346.

29-NOV-1996; 96US-0757669.

25-JUL-1996; 96US-0686243.

25-JUL-1996; 96US-0686349.

25-JUL-1996; 96US-0686350.

25-JUL-1996; 96US-0687253.

25-JUL-1996; 96US-0688814.

05-SEP-1996; 96US-0708678.

10-OCT-1996; 96US-0728323.

13-NOV-1996; 96US-0747887.

13-NOV-1996; 96US-0748640.

(UYCO ) UNIV COLUMBIA NEW YORK.

PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;

WPI; 1998-130615/12.

New nucleic acid encoding Kaposi's sarcoma associated herpes virus  
 proteins - useful for, e.g. detecting levels of HHV8 in, and  
 preparation of vaccines for treatment of, HIV patients

Example 2; Page 135-203; 230pp; English.

This sequence represents the long unique region and terminal repeat of  
 the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known  
 as human herpes virus 8 (HHV8). This sequence contains the DNAs of the  
 invention which encode KSHV polypeptides selected from: (a) viral  
 macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);  
 (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;  
 (d) capsid protein IV encoded by ORF65; and (e) immediate early protein  
 encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded  
 by it, and antibodies (Ab) specific for the proteins are useful for  
 detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body  
 fluids or tissue samples. HHV8 infections can be treated with antisense  
 or triplex forming molecules or agents that bind specifically to the  
 protein. Ab may be used for prophylaxis or treatment of HHV8 infection,  
 while the protein can be used in protective vaccines. Ab may also be used  
 to differentiate between lymphomas, and HHV8 may be implicated in many  
 other lymphoproliferative diseases such as lymphomas, leukaemia,  
 splenomegaly and mycosis fungoides. Cells and animals containing the  
 nucleic acid are useful for drug screening. HHV8-derived peptides can be  
 used as targets for antiviral drugs, e.g. dihydrofolate reductase gene  
 can be inhibited with methotrexate. These can also be used to determine  
 the immune status of a patient infected with HIV. HHV8 derived protein  
 viral MIP III may be used as an anti-inflammatory agent for,  
 e.g. treating rheumatoid arthritis. This sequence is stated as containing  
 81 open reading frames.

Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

Query Match

Best Local Similarity 1.7%; Score 18; DB 19; Length 137507;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 443 caccagagcaccacgaag 460

Db 115914 caccagagcaccacgaag 115931  
 RESULT 14  
 Q79912/c

ID Q79912 standard; DNA; 831 BP.

XX Q79912;

AC Q79912;

XX 02-OCT-1995 (first entry)

XX Aromatic dihydrodiol dihydrogenase gene.

XX Dihydrodiol; catechol; microbial fermentation; E.coli; aromatic;

KW deoxygenase; hydrocarbon; polymerisation inhibitor; synthetic resin;

KW drug synthesis; ds.

XX Pseudomonas fluorescens SD805.

XX JP06343474-A.

XX 20-DEC-1994.

XX 10-JUN-1993; 33JP-0165163.

XX 10-JUN-1993; 33JP-0165163.

XX (SHOW ) SHOWA DENKO KK.

XX WPI; 1995-069305/10.

DR P-PSDB; R66734.

XX Prepn of an aromatic cyclic dihydrodiol cpd and a catechol cpd

PT - using microbial cells transformed by DNA coding for enzyme

PT protein sequences

XX Claim 3; Page 7; 13pp; Japanese.

XX The genes (Q79907-12) encode enzymes involved in the synthesis of

CC dihydrodiol and catechol compounds by microbial fermentation in E.coli.

CC The sequence presented here encodes an aromatic dihydrodiol dihydrogenase

CC of 276 amino acids. The enzymes encoded are used in methods for the

CC synthesis of dihydrodiol compounds and catechol compounds from an

CC aromatic hydrocarbon at normal temperature and pressure. The aromatic

CC dihydrodiol can be used as a raw material for engineering plastics and

CC the catechol compound is used as a polymerisation inhibitor of synthetic

CC resin and as an intermediate in drug synthesis.

XX Sequence 831 BP; 185 A; 191 C; 253 G; 202 T; 0 other;

SQ

Query Match

Best Local Similarity 1.6%; Score 17; DB 16; Length 831;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 375 aggtctctcaaatccg 391

Db 582 AGGTCTCTCAAAATCCG 566

RESULT 15

ID N40050

XX N40050 standard; DNA; 1569 BP.

AC N40050;

XX 02-FEB-1992 (first entry)

DE Sequence of 'a'-factor structural gene.

XX Yeast 'a'-factor; promoter; yeast expression vector;

KW secretion vector; ss.

```

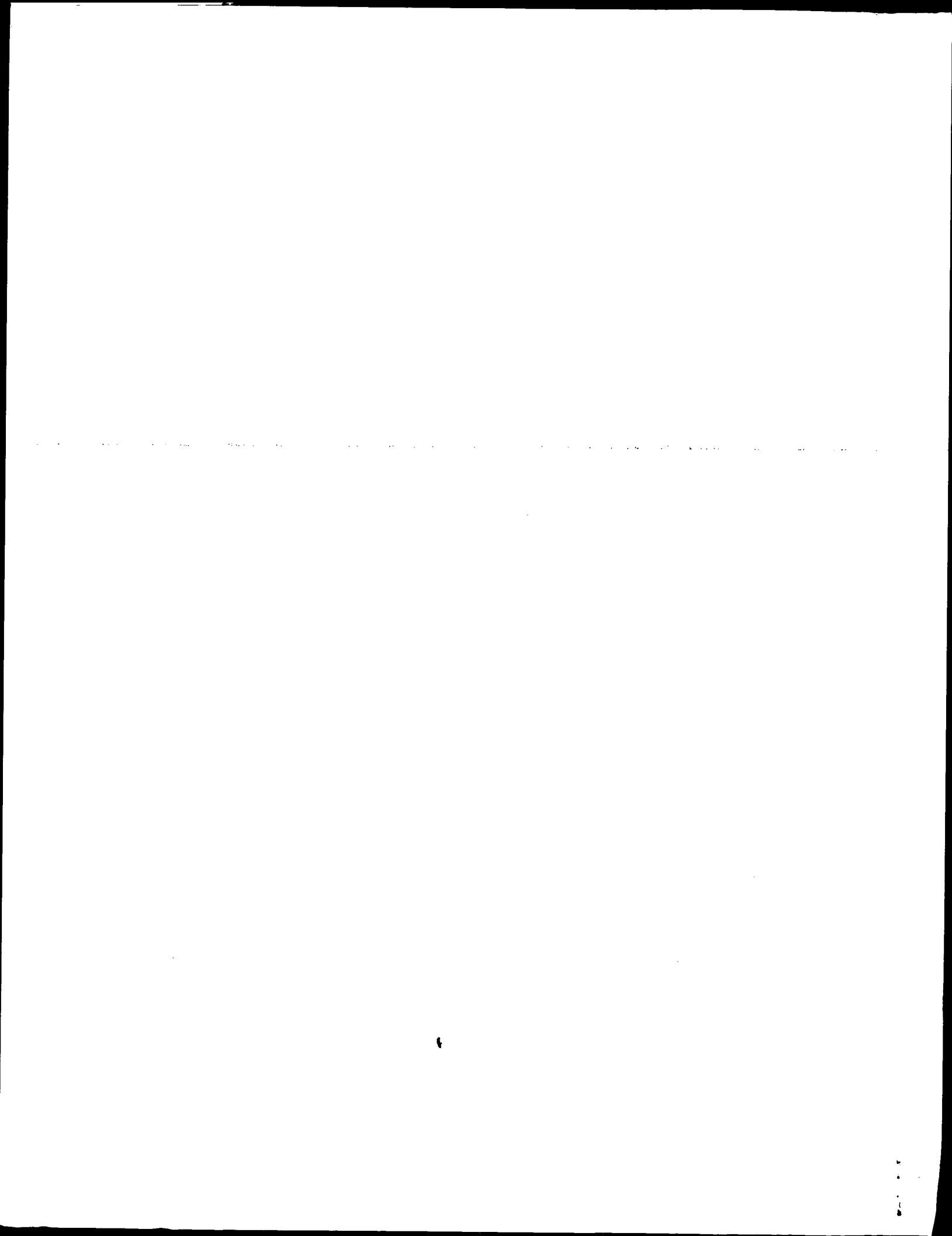
XX OS Yeast.
XX Key Location/Qualifiers
XX Key 626..736
XX FT CDS /*tag= a
XX FT
XX XX
XX PN EP123289-A.
XX XX 31-OCT-1984.
XX XX 19-APR-1984; 84EP-0104445.
XX XX 26-APR-1983; 83US-0488857.
XX XX (CHIR-) CHIRON CORP.
XX XX
XX PI Brake AJ;
XX XX WPI; 1984-271262/44.
XX DR P-PSDB; P40037.
XX XX
XX PT DNA construct contg. yeast leader sequence of A-factor - useful
XX PT for efficient expression of polypeptide(s) in yeast
XX PS Example; Page 16; 26pp; English.
XX XX
XX CC The inventors claim a DNA construct comprising the yeast leader
XX CC sequence of 'a'-factor including processing signals and a
XX CC heterologous gene in reading frame with the leader sequence and
XX CC processing signals. The pref. promoter is an 'a'-factor promoter.
XX XX
XX SQ Sequence 1569 BP; 485 A; 308 C; 308 G; 468 T; 0 other;

Query Match 1.6%; Score 17; DB 5; Length 1569;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 aagctgtctctgtgtgc 182
   |||||
Db 1063 aagctgtctctgtgtgc 1079

```

Search completed: November 4, 2000, 03:40:56  
Job time: 11320 sec



us-09-101-423a-6.olig.rni

Mon Nov 6 10:14:41 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 03:38:09 ; Search time 140.17 Seconds  
(without alignments)  
1176.095 Million cell updates/sec

Title: US-09-101-423A-6  
Perfect score: 1090  
Sequence: 1 ttgtctgtgttacttgt.....cctagcacagtgcattccaca 1090

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 262060 seqs, 75620727 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/PCRTUS\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.7	1374	US-08-929-738-1	Sequence 1, Appl
C 2	19	1.7	1614	US-08-929-738-2	Sequence 2, Appl
C 3	18	1.7	32207	US-08-770-379-20	Sequence 20, Appl
C 4	18	1.7	51259	US-08-781-891-209	Sequence 209, Appl
C 5	17	1.6	34	US-08-482-577B-21	Sequence 21, Appl
C 6	17	1.6	2244	US-08-476-519-10	Sequence 10, Appl
C 7	17	1.6	2244	US-08-476-519-1	Sequence 1, Appl
C 8	17	1.6	2334	US-08-476-519-1	Sequence 1, Appl
C 9	17	1.6	2334	US-08-476-519-1	Sequence 1, Appl
C 10	16	1.5	327	US-08-154-916-1	Sequence 1, Appl
C 11	16	1.5	1066	US-08-139-424-1	Sequence 1, Appl
C 12	16	1.5	1095	US-08-448-744-6	Sequence 6, Appl
C 13	16	1.5	1128	US-08-230-047-4	Sequence 4, Appl
C 14	16	1.5	1168	US-08-282-581-3	Sequence 3, Appl
C 15	16	1.5	1388	US-08-550-544-3	Sequence 3, Appl
C 16	16	1.5	1642	US-08-665-037-1	Sequence 1, Appl
C 17	16	1.5	1642	US-08-666-067-1	Sequence 1, Appl
C 18	16	1.5	1642	US-08-732-870-1	Sequence 1, Appl
C 19	16	1.5	1642	US-08-299-849B-25	Sequence 25, Appl
C 20	16	1.5	2099	US-08-142-368A-25	Sequence 25, Appl
C 21	16	1.5	2099	US-08-967-727-25	Sequence 25, Appl
C 22	16	1.5	2099	US-08-299-849B-24	Sequence 24, Appl
C 23	16	1.5	2150	US-08-142-368A-24	Sequence 24, Appl
C 24	16	1.5	2150	US-08-967-727-24	Sequence 24, Appl
C 25	16	1.5	2150	US-08-967-727-24	Sequence 24, Appl
C 26	16	1.5	2431	US-08-985-526-35	Sequence 35, Appl

27	16	1.5	3918	5	US-08-836-329-1	Sequence 1, Appl
C 28	16	1.5	5406	1	US-07-813-593-3	Sequence 3, Appl
C 29	16	1.5	5406	1	US-07-977-451-5	Sequence 5, Appl
C 30	16	1.5	5406	1	US-07-946-507-3	Sequence 3, Appl
C 31	16	1.5	5406	1	US-08-252-517-5	Sequence 5, Appl
C 32	16	1.5	5406	1	US-07-906-397A-5	Sequence 5, Appl
C 33	16	1.5	5406	2	US-08-601-891-5	Sequence 5, Appl
C 34	16	1.5	5406	4	US-09-021-324-5	Sequence 5, Appl
C 35	16	1.5	5406	6	PCT-US92-02750-7	Sequence 7, Appl
C 36	16	1.5	5406	6	PCT-US92-03401-5	Sequence 5, Appl
C 37	16	1.5	5406	6	PCT-US92-09893-5	Sequence 5, Appl
C 38	16	1.5	5470	3	US-08-443-861-1	Sequence 1, Appl
C 39	16	1.5	5470	3	US-09-045-201A-1	Sequence 1, Appl
C 40	16	1.5	10607	2	US-08-078-090-3	Sequence 3, Appl
C 41	16	1.5	35060	5	US-08-814-095-7	Sequence 7, Appl
C 42	15	1.4	24	4	US-08-859-998-549	Sequence 549, Appl
C 43	15	1.4	31	5	US-08-115-753-23	Sequence 23, Appl
C 44	15	1.4	36	1	US-08-147-000B-14	Sequence 14, Appl
C 45	15	1.4	80	5	US-08-933-983-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1  
US-08-929-738-1/c  
; Sequence 1, Application US/08929738  
; Patent No. 6020149  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Anti-microbial Agents and Screening Method  
; NUMBER OF SEQUENCES: 2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IEM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/929,738  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB94/000649  
; FILING DATE:  
; APPLICATION NUMBER: GB 9306416.0  
; FILING DATE: 27-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9401532.8  
; FILING DATE: 27-JAN-1994  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1374 base pairs  
; TYPE: nucleic acid  
; TOPOLOGY: double  
; STRANDEDNESS: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NC  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: DH5 alpha  
; FEATURE:  
; NAME/KEY: misc\_signal  
; LOCATION: 1..3  
; OTHER INFORMATION: /function="Altered transcription  
; OTHER INFORMATION: Initiation codon"  
; OTHER INFORMATION: /label= ATGI  
; FEATURE:  
; NAME/KEY: misc\_signal  
; LOCATION: 75..77  
; OTHER INFORMATION: /function="Formerly thought to be  
; OTHER INFORMATION: translation initiation codon"  
; OTHER INFORMATION: /label= ATG2

US-08-929-738-1

Query Match 1.7%; Score 19; DB 5; Length 1374;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 651 cagaggaatttgccttg 669  
|||||  
Db 693 CAGAGGAATTTGCCTTG 675

RESULT 2

US-08-929-738-2/c  
; Sequence 2, Application US/08929738  
; Patent No. 6020149  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Anti-microbial Agents and Screening Method  
; NUMBER OF SEQUENCES: 2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/929,738  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB94/000649  
; FILING DATE:  
; APPLICATION NUMBER: GB 9306416.0  
; FILING DATE: 27-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9401532.8  
; FILING DATE: 27-JAN-1994  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1614 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: DH5 alpha  
; FEATURE:  
; NAME/KEY: misc.signal  
; LOCATION: 130..132  
; OTHER INFORMATION: /function= "Altered translation  
; OTHER INFORMATION: Initiation codon"  
; OTHER INFORMATION: /label= ATG1  
; FEATURE:  
; NAME/KEY: misc.signal  
; LOCATION: 205..207  
; OTHER INFORMATION: /function= "Formerly thought to be  
; OTHER INFORMATION: translation initiation codon"  
; OTHER INFORMATION: /label= ATG2

US-08-929-738-2

Query Match 1.7%; Score 19; DB 5; Length 1614;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 651 cagaggaatttgccttg 669  
|||||  
Db 822 CAGAGGAATTTGCCTTG 804

RESULT 3

US-08-770-379-20  
; Sequence 20, Application US/08770379  
; Patent No. 5849564  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohanzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Edelman, Isidore S.  
; APPLICANT: Moore, Patrick S.  
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
; HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/770,379  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 52342  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32207 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-770-379-20

Query Match 1.7%; Score 18; DB 3; Length 32207;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 443 caccagagggcaccgaag 460  
|||||  
Db 10614 CACCAGAGGCACCAAG 10631

RESULT 4

US-08-781-891-209  
; Sequence 209, Application US/08781891  
; Patent No. 6090620  
; GENERAL INFORMATION:  
; APPLICANT: Fu, Ying-Hui  
; APPLICANT: Yu, Chang-En  
; APPLICANT: Oshima, Junko  
; APPLICANT: Mulligan, John T.  
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
; WERNER'S SYNDROME  
; NUMBER OF SEQUENCES: 209  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington

US-08-781-891-209

```

; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-781-891-209

```

Query Match 1.7%; Score 18; DB 5; Length 51259;  
 Best Local Similarity 100.0%; Pred. No. 7.3;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 ctgtcagatgtgtagctg 146  
 |||||  
 DB 50327 CTGTCAGATGTGTAGCTG 50344

```

RESULT 5
US-08-482-577B-21/c
; Sequence 21, Application US/08482577B
; Patent No. 5807713
; GENERAL INFORMATION:
; APPLICANT: HOTTEN, GERTRUD
; APPLICANT: NEIDHARDT, HELGE
; APPLICANT: BECHTOLD, ROLF
; APPLICANT: FOHL, JENS
; TITLE OF INVENTION: DNA SEQUENCES ENCODING NOVEL
; TITLE OF INVENTION: GROWTH/DIFFERENTIATION FACTORS
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAI, MARMELSTEIN, MURRAY, AND ORAM
; STREET: 655 FIFTEENTH STREET, N.W., G STREET LOBBY,
; STREET: SUITE 330
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,577B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KLESNER, SHARON
; REGISTRATION NUMBER: 36,335
; REFERENCE/DOCKET NUMBER: P564-5010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 21:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-482-577B-21

```

Query Match 1.6%; Score 17; DB 2; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 644 caggtcccaaggaatt 660  
 |||||  
 DB 24 CAGGTCCCAAGGAATT 8

```

RESULT 6
US-08-476-519-10
; Sequence 10, Application US/08476519
; Patent No. 5750876
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Krich, Bradley M.
; TITLE OF INVENTION: No. 5750876el Isoamylase Gene, Compositions
; TITLE OF INVENTION: Containing It and Methods of Using Isoamylases
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace L. Bonner, Monsanto Company, B84F
; STREET: 700 Chesterfield Parkway No. 5750876th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,519
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/281902
; FILING DATE: 28-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bonner, Grace L.
; REGISTRATION NUMBER: 32,963
; REFERENCE/DOCKET NUMBER: 38-21(13577)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)537-7286
; TELEFAX: (314)537-6047
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2241
; US-08-476-519-10

```

Query Match 1.6%; Score 17; DB 2; Length 2244;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1040 c9gcaccggactg9cg 1056

```
Db 1646 CGCGCACCAGCTGGCG 1662
|||||
RESULT 7
PCT-US95-09323-10
; Sequence 10, Application PC/TUS9509323
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Novel Isoamylase Gene, Compositions
; TITLE OF INVENTION: Containing It and Methods of Using Isoamylases
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09323
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/281902
; FILING DATE: 28-JUL-1994
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2241
PCT-US95-09323-10

Query Match 1.6%; Score 17; DB 6; Length 2244;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1040 CGCGCACCAGCTGGCG 1056
|||||
Db 1646 CGCGCACCAGCTGGCG 1662

RESULT 8
US-08-476-519-1
; Sequence 1, Application US/08476519
; Patent No. 5750876
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Krohn, Bradley M.
; TITLE OF INVENTION: No. 5750876el Isoamylase Gene, Compositions
; TITLE OF INVENTION: Containing It and Methods of Using Isoamylases
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F
; STREET: 700 Chesterfield Parkway No. 5750876th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,519
; FILING DATE:
; CLASSIFICATION: 800

PCT-US95-09323-1
; Sequence 1, Application PC/TUS9509323
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Novel Isoamylase Gene, Compositions
; TITLE OF INVENTION: Containing It and Methods of Using Isoamylases
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09323
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/281902
; FILING DATE: 28-JUL-1994
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-09323-1

Query Match 1.6%; Score 17; DB 6; Length 2334;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1040 CGCGCACCAGCTGGCG 1056
|||||
Db 1739 CGCGCACCAGCTGGCG 1755

RESULT 9
PCT-US95-09323-1
; Sequence 1, Application PC/TUS9509323
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Novel Isoamylase Gene, Compositions
; TITLE OF INVENTION: Containing It and Methods of Using Isoamylases
; NUMBER OF SEQUENCES: 11
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09323
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/281902
; FILING DATE: 28-JUL-1994
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US95-09323-1

Query Match 1.6%; Score 17; DB 6; Length 2334;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1040 CGCGCACCAGCTGGCG 1056
|||||
Db 1739 CGCGCACCAGCTGGCG 1755

RESULT 10
US-08-741-406-6
; Sequence 6, Application US/08741406
; Patent No. 5721118
; GENERAL INFORMATION:
```



APPLICANT: Scheffler, Immo E.  
TITLE OF INVENTION: Mammalian Artificial Chromosomes and  
METHODS OF USING SAME  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/741,406  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/550,717  
FILING DATE: 31-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-UD 2317  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 327 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-741-406-5

Query Match 1.5%; Score 16; DB 2; Length 327;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 484 ggcaagtggtggcaga 499  
|||||  
DB 137 GGCAAGGTGGGCAGA 152

RESULT 11  
US-08-154-916-1  
Sequence 1, Application US/08154916  
Patent No. 5468612  
GENERAL INFORMATION:  
APPLICANT: Cohen, Edward Hirsch  
APPLICANT: Landgraf, Bryan E.  
TITLE OF INVENTION: 9804 Gene and Method of Use  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/154,916  
FILING DATE: NO. 5468612ember 19, 1993

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 06180/003001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1066  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-154-916-1

Query Match 1.5%; Score 16; DB 1; Length 1066;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 534 tgccctcacttctggg 549  
|||||  
DB 634 TGCCCTCACTCTGGG 649

RESULT 12  
US-09-139-424-1  
Sequence 1, Application US/09139424  
Patent No. 5985832  
GENERAL INFORMATION:  
APPLICANT: Rodman, G. David  
APPLICANT: Reddy, Sakamuri V.  
APPLICANT: Chci, Sun-Jin  
TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE FOR  
TITLE OF INVENTION: OSTEOCLAST INHIBITORY FACTOR  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/139,424  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/772,441  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: UTSK:295  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000  
TELEFAX: 512/414-7577  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1095 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

FEATURE:  
NAME/KEY: CDS  
LOCATION: 95..487  
US-09-139-424-1

Query Match 1.5%; Score 16; DB 4; Length 1095;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 534 tgcctcacttctggg 549  
Db 674 tgcctcacttctggg 689

## RESULT 13

US-08-448-744-6  
Sequence 6, Application US/08448744  
Patent No. 5616484

## GENERAL INFORMATION:

APPLICANT: XU, Shuang-Yong  
TITLE OF INVENTION: Cloning And Expression of The Apali  
TITLE OF INVENTION: Restriction Endonuclease  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: New England Biolabs, Inc.  
STREET: 32 Tozer Road  
CITY: Beverly  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 01915

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/448,744  
FILING DATE: 24-May-1995  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Williams, Gregory D.  
REGISTRATION NUMBER: 30901  
REFERENCE/DOCKET NUMBER: NEB-114

TELEPHONE: (508) 927-5054  
TELEFAX: (508) 927-1705

## INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1128 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-448-744-6

Query Match 1.5%; Score 16; DB 1; Length 1128;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 995 atccttctgcagcca 1010  
Db 541 ATCCTTCTGCAGCCA 556

## RESULT 14

US-08-230-047-4/c  
Sequence 4, Application US/08230047  
Patent No. 5541109

## GENERAL INFORMATION:

APPLICANT: Searfoss III, George H.  
APPLICANT: Ivashchenko, Yuri D.

APPLICANT: Jays, Michael C.  
TITLE OF INVENTION: EXPRESSION CLONING OF C-SRC SH3 BINDING  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Road, 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: System 7.1  
SOFTWARE: Word 5.0 (PatentIn)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/230,047  
FILING DATE: 19-APR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Goodman, Rosanne  
REGISTRATION NUMBER: 32,534  
REFERENCE/DOCKET NUMBER: A1465-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 454-3817  
TELEFAX: (610) 454-3808  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1168 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS3  
LOCATION: 1..1110  
US-08-230-047-4

Query Match 1.5%; Score 16; DB 1; Length 1168;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 tgggtgcagggagctg 367  
Db 27 TGGGTGCAGGGAGCTG 12

## RESULT 15

US-08-282-581-3/c  
Sequence 3, Application US/08282581  
Patent No. 5670349

## GENERAL INFORMATION:

APPLICANT: Cramer, Carole L.  
APPLICANT: Weissenborn, Deborah L.

TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM AND

TITLE OF INVENTION: POST-HARVEST PRODUCTION OF GENE PRODUCTS

TITLE OF INVENTION: IN PLANTS AND PLANT CELL CULTURES

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

us-09-101-423a-6.olig.rni

Mon Nov 6 10:14:41 2000

```

APPLICATION NUMBER: US/08/282,581
FILING DATE: 29-JUL-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18872
REFERENCE/DOCKET NUMBER: 7956-009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1388 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-282-581-3

```

```

Query Match      1.5%; Score 16; DB 1; Length 1388;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 931 atttctatttcttt 946
    |||||
Db 643 ATTCTATTCTTTT 528

```

```

Search completed: November 4, 2000, 03:38:47
Job time: 12317 sec

```



us-09-101-423a-6.olig.rst

Mon Nov 6 10:14:42 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 01:08:54 ; Search time 3299.68 Seconds  
(without alignments)  
2042.398 Million cell updates/sec

Title: US-09-101-423A-6  
Perfect score: 1090  
Sequence: 1 ttgtctcgggttactgtt.....cctagcacagtgcaccca 1090

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 7189864 seqs, 3091403243 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries

Database : EST:  
1: gb\_estl1:  
2: gb\_est2:  
3: gb\_est3:  
4: gb\_est4:  
5: gb\_est5:  
6: gb\_est6:  
7: gb\_est7:  
8: gb\_est8:  
9: gb\_est9:  
10: gb\_est10:  
11: gb\_est11:  
12: gb\_est12:  
13: gb\_est13:  
14: gb\_est14:  
15: gb\_est15:  
16: gb\_est16:  
17: gb\_est17:  
18: gb\_est18:  
19: gb\_est19:  
20: gb\_est20:  
21: gb\_est21:  
22: gb\_est22:  
23: gb\_est23:  
24: gb\_est24:  
25: gb\_est25:  
26: gb\_est26:  
27: gb\_est27:  
28: gb\_est28:  
29: gb\_est29:  
30: gb\_est30:  
31: gb\_est31:  
32: gb\_est32:  
33: gb\_est33:  
34: gb\_est34:  
35: gb\_est35:  
36: gb\_est36:  
37: gb\_est37:  
38: gb\_est38:  
39: gb\_est39:  
40: gb\_est40:  
41: em\_estba:  
42: em\_estfun:  
43: em\_esthum1:

44: em\_esthum2:  
45: em\_esthum3:  
46: em\_esthum4:  
47: em\_esthum5:  
48: em\_esthum6:  
49: em\_esthum7:  
50: em\_esthum8:  
51: em\_esthum9:  
52: em\_esthum10:  
53: em\_esthum11:  
54: em\_esthum12:  
55: em\_esthum13:  
56: em\_esthum14:  
57: em\_esthum15:  
58: em\_esthum16:  
59: em\_esthum17:  
60: em\_esthum18:  
61: em\_esthum19:  
62: em\_esthum20:  
63: em\_estin1:  
64: em\_estin2:  
65: em\_estin3:  
66: em\_estin4:  
67: em\_estom:  
68: em\_estov1:  
69: em\_estov2:  
70: em\_estpl1:  
71: em\_estpl2:  
72: em\_estpl3:  
73: em\_estpl4:  
74: em\_estpl5:  
75: em\_estrol:  
76: em\_estro2:  
77: em\_estro3:  
78: em\_estro4:  
79: em\_estro5:  
80: em\_estro6:  
81: em\_estro7:  
82: em\_estro8:  
83: em\_estro9:  
84: em\_estrol0:  
85: em\_estrol1:  
86: em\_estrol2:  
87: em\_estrol3:  
88: gb\_gss1:  
89: gb\_gss2:  
90: gb\_gss3:  
91: gb\_gss4:  
92: em\_gss1:  
93: em\_gss2:  
94: em\_gss3:  
95: em\_gss4:  
96: gb\_gss5:  
97: gb\_gss6:  
98: gb\_gss7:  
99: gb\_gss8:  
100: gb\_gss9:  
101: em\_gss5:  
102: em\_gss6:  
103: em\_gss7:  
104: em\_gss8:  
105: em\_gss9:  
106: em\_gss10:  
107: em\_gss11:  
108: gb\_gss10:  
109: gb\_gss11:  
110: em\_gss12:  
111: gb\_gss12:  
112: gb\_gss13:  
113: gb\_gss14:  
114: gb\_gss15:  
115: gb\_gss16:  
116: gb\_gss17:

117: gb\_gss18:\*  
 118: gb\_gss19:\*  
 119: em\_gss13:\*  
 120: gb\_gss20:\*  
 121: gb\_gss21:\*  
 122: gb\_gss22:\*  
 123: gb\_gss23:\*  
 124: gb\_gss24:\*  
 125: em\_gss14:\*  
 126: em\_gss15:\*  
 127: em\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	28	2.6	493	9	AI237080	EST233642
2	26	2.4	457	9	AI176946	EST220552
3	24	2.2	254	11	AI555421	EST220552
4	23	2.1	368	37	H33201	UI-R-C2p-
5	22	2.0	486	118	B49308	EST108975 R
6	21	1.9	505	25	AW921512	RPC111-5N20
7	21	1.9	612	24	AW773148	EST352816
8	20	1.8	495	11	AI599832	EST251535
9	20	1.8	594	117	AZ228599	RPC11-23-5
10	20	1.8	624	114	AZ028284	RPC1-23-3
11	20	1.8	654	117	AZ229370	RPC1-23-5
12	20	1.8	812	15	AU080850	AU080850
13	20	1.8	1101	121	CNS000EFO	AU080850
14	19	1.7	234	35	BE344801	AI069133 Drosophila
15	19	1.7	321	100	AQ641537	BE344801 946029B02
16	19	1.7	322	32	B544069	AQ641537 RPC133-Dp
17	19	1.7	344	39	R82709	BS44069 BS44069
18	19	1.7	377	111	AQ89364	R82709 YJ20F01.r1
19	19	1.7	442	33	BE096067	AQ89364 nbe0035K
20	19	1.7	459	23	AW577127	BE096067 UI-R-C4-a
21	19	1.7	459	23	AW601412	AW577127 MR4-BT036
22	19	1.7	525	10	AI324640	AW601412 MR4-BT035
23	19	1.7	535	11	AI552807	AI324640 ms60602.x
24	19	1.7	545	11	AQ823463	AI552807 mm27d12.y
25	19	1.7	555	13	AI796210	AQ823463 HS_3217.A
26	19	1.7	572	108	AQ705125	AI796210 wh44809.x
27	19	1.7	604	21	AW360847	AQ705125 HS_5532.B
28	19	1.7	842	108	AQ747039	AW360847 PM3-CT024
29	19	1.7	1075	112	AQ896553	AQ747039 HS_5538.A
30	18	1.7	111	100	AQ634901	AQ896553 HS_3148.B
31	18	1.7	129	7	AI001708	AQ634901 RPC1-11-4
32	18	1.7	168	35	BE433304	AI001708 EST0230 T
33	18	1.7	200	16	AV126693	BE433304 EST399833
34	18	1.7	205	15	AV097634	AV126693 AV126693
35	18	1.7	219	11	AI537511	AV097634 AV097634
36	18	1.7	220	22	AW462114	AI537511 tp09h11.x
37	18	1.7	224	18	AV338937	AW462114 BP230008A
38	18	1.7	255	22	BB496939	AV338937 AV338937
39	18	1.7	257	22	AW512872	BB496939 BB496939
40	18	1.7	258	7	AA943175	AW512872 x004a12.x
41	18	1.7	258	37	H32274	AA943175 EST198674
42	18	1.7	272	11	AI556727	H32274 EST107204 R
43	18	1.7	283	38	R59578	AI556727 UI-R-C2p-
44	18	1.7	284	10	AI354851	R59578 yho2a06.sl
45	18	1.7	287	31	BB456154	AI354851 qt76a12.x
						BB456154 BB456154

## ALIGNMENTS

RESULT 1  
 AI237080

LOCUS AI237080 493 bp mRNA EST 31-JAN-1999  
 DEFINITION EST233642 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone  
 ROVD045 3' end, mRNA sequence.  
 ACCESSION AI237080  
 VERSION AI237080.1 GI:3830586  
 KEYWORDS EST.  
 SOURCE Rattus sp.  
 ORGANISM Rattus sp.  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 493)  
 AUTHORS Lee,N.H.; Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.  
 TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Lee, NH  
 ATCC The Institute for Genomic Research  
 9712, Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208  
 Email: niles@tigr.org  
 Seq primer: M13-21.  
 FEATURES  
 source location/Qualifiers  
 1. 493  
 /organism="Rattus sp."  
 /db\_xref="ATCC (inhost):2042239"  
 /db\_xref="taxon:10118"  
 /clone="ROVD045"  
 /note\_lib="Normalized rat ovary, Bento Soares"  
 /note="Organ: ovary; Vector: pT7T3pac; Site\_1: EcoRI; Site\_2: NotI"  
 BASE COUNT 86 a 110 c 133 g 163 t 1 others  
 ORIGIN  
 Query Match 2.6%; Score 28; DB 9; Length 493;  
 Best Local Similarity 100.0%; Pred. No. 0.00081;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 683 caccagcgaggtcacttgagcagcaaaa 710  
 ||||||||||||||||||||||||||||  
 Db 181 CACCAGCAGGTCACCTGGAGCAGAAAA 208  
 RESULT 2  
 AI176946  
 LOCUS AI176946 457 bp mRNA EST 08-JAN-1999  
 DEFINITION EST220552 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone  
 ROVD92 3' end, mRNA sequence.  
 ACCESSION AI176946  
 VERSION AI176946.1 GI:4134938  
 KEYWORDS EST.  
 SOURCE Rattus sp.  
 ORGANISM Rattus sp.  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE 1 (bases 1 to 457)  
 AUTHORS Lee,N.H.; Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.  
 TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index  
 JOURNAL Unpublished (1998)  
 COMMENT On Oct 8, 1998 this sequence version replaced gi:3727584.  
 Contact: Lee, NH  
 ATCC The Institute for Genomic Research  
 9712, Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208

-- MON NOV 6 10:14:42 2000

Email: nhlee@tigr.org  
Seq primer: M13-21

FEATURES  
source Location/Qualifiers

1. .457  
/organism="Rattus sp."  
/db\_xref="taxon:10118"  
/clone="ROV392"  
/note="Normalized rat ovary, Bento Soares"  
/note="Organ: ovary; Vector: pMT33pac; Site\_1: EcoRI;  
Site\_2: NotI"

BASE COUNT 79 a 115 c 120 g 143 t

ORIGIN

Query Match 2.4%; Score 26; DB 9; Length 457;  
Best Local Similarity 100.0%; Pred. No. 0.0096;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 ttctctgttttttttcagcagttta 110  
|||||

DB 150 TTTCTGTTTCTTTTCAGCCAGTTA 175  
|||||

RESULT 3

AI555421 254 bp mRNA EST 23-MAR-1999  
LOCUS UI-R-C2p-qp-f-08-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone  
DEFINITION UI-R-C2p-qp-f-08-0-UI 3', mRNA sequence.

ACCESSION AI555421

VERSION AI555421.1 GI:4487784

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 254)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 9704477

COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu  
cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone  
distribution: clones will be available through Research Genetics  
(www.resgen.com) The following repetitive elements were found in  
this cDNA sequence: 8-246, >L1\_RN#LINE/L1  
Seq primer: M13 Forward.

FEATURES Location/Qualifiers

source 1. .254  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10118"  
/clone="UI-R-C2p-qp-f-08-0-UI"  
/clone\_lib="UI-R-C2p"  
/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pMT33D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-C2p  
library is a subtracted library derived from the UI-R-C1  
library, which is a subtracted library derived from the  
UI-R-C0 library. The UI-R-C0 library consisted of a  
mixture of individually tagged normalized libraries  
constructed from rat placenta, adult lung, brain, liver,  
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day  
embryo. The tag is a string of 3-5 nucleotides present  
between the Not I site and the oligo-dT track which allows

identification of the library of origin of a clone within  
the mixture. The subtracted library (UI-R-C2p) was  
constructed as follows: PCR amplified cDNA inserts from  
UI-R-C1 clones from which 3' ESTs had been derived was  
used as a driver in a hybridization with the UI-R-C1  
library in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library) was  
purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (life technologies) to generate the  
UI-R-C2p library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research 6:  
791-806, 1996).

BASE COUNT 32 a 61 c 61 g 100 t

ORIGIN

Query Match 2.2%; Score 24; DB 11; Length 254;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 ttctgttttttttcagcagttta 110  
|||||

DB 137 TTCTGTTTCTTTTCAGCCAGTTA 160  
|||||

RESULT 4

H33201

LOCUS

DEFINITION

EST108975 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA

clone RFNAJ24 3' end, mRNA sequence.

ACCESSION H33201

VERSION H33201.1 GI:978618

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 368)

AUTHORS Lee,N.H.; Weinstein,K.G.; Kirkness,E.F.; Earle-Rughes,J.A.; Fuldner

,R.A.; Marmaras,S.; Glodex,A.; Gocayne,J.D.; Adams,M.D.; Kerlavage

,A.R.; Fraser,C.M. and Venter,J.C.

TITLE Comparative expressed-sequence-tag analysis of differential gene

expression profiles in PC-12 cells before and after nerve growth

factor treatment

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)

MEDLINE 95396786

COMMENT Other ESTs: EST108976

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

For clone availability please contact the TIGR Database

(tdbinfo@tigr.org)

Seq primer: M13 - 21.

FEATURES Location/Qualifiers

source 1. .368

/organism="Rattus sp."

/db\_xref="Arcc (Inhost):2003480"

/db\_xref="taxon:10118"

/clone="RFNAJ24"

/clone\_lib="Rat PC-12 cells, NGF-treated (9 days)"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; poly(A)+ RNA was purified from 9-day NGF treated

PC12 cells. cDNA was constructed using an oligo-dT primer

and directionally cloned using the Lambda ZAP II Vector

Kit by Stratagene"

BASE COUNT 80 a 77 c 90 g 119 t 2 others

ORIGIN

```

Query Match      2.18; Score 23; DB 37; Length 368;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 gaaatgtggcgagagtagtct 291
|||||
Db 52 GAAATGTGGCAGAGTAGTCT 74

RESULT 5
B49308/c
LOCUS
DEFINITION RPC111-5N20-TV RPCI-11 Homo sapiens genomic clone RPCI-11-5N20, DNA
sequence.
ACCESSION B49308
VERSION B49308.1 GI:2601545
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 486)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPC111-5N20.TP
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamads@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: 17
Class: BAC ends.
FEATURES             Location/Qualifiers
     source
     1..486
     /organism="Homo sapiens"
     /db_xref="GDB:7501867"
     /db_xref="taxon:9606"
     /clone="RPCI-11-5N20"
     /clone_lib="RPCI-11"
     /sex="Male"
     /cell_type="Lymphocytes"
     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC library"
BASE COUNT      148 a 115 c 93 g 130 t
ORIGIN

Query Match      2.08; Score 22; DB 118; Length 486;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 924 cagctccattctatttcttt 945
|||||
Db 138 CAGCTCCATTCTATTTCCTTT 117

RESULT 6
AW921512
LOCUS
DEFINITION EST352816 Rat gene index, normalized rat, norvegicus, Bento Soares

```

```

Rattus norvegicus CDNA clone RGI1A14 5' end, mRNA sequence.
ACCESSION AW921512
VERSION AW921512.1 GI:8087336
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 505)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
ATCC The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information
Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source
     1..505
     /organism="Rattus norvegicus"
     /db_xref="taxon:10116"
     /clone="RGI1A14"
     /clone_lib="Rat gene index, normalized rat, norvegicus,
Bento Soares"
     /tissue_type="mix - brain, ovary, placenta, kidney, lung,
liver, embryo, heart, muscle, spleen"
     /lab_host="SOLR"
     /note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; Estimated insert size approx.1 kb"
BASE COUNT      72 a 123 c 149 g 161 t
ORIGIN

Query Match      1.9%; Score 21; DB 25; Length 505;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 ttctctgtttcttctcagcc 105
|||||
Db 14 TTTTCTGTCTTCTTCAGCC 34

RESULT 7
AW773148
LOCUS
DEFINITION AW773148. 612 bp mRNA
ACCESSION EST00054 DT40subNB Gallus gallus cDNA clone 1H9, mRNA sequence.
VERSION AW773148.1 GI:7709716
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 612)
AUTHORS Nelman,P.E. and Ruddell,A.R.
TITLE Complementary DNA sequencing: Bursal lymphoma cell line, DT-40
subtracted by normal bursa
JOURNAL Unpublished (2000)
COMMENT Contact: Nelman, P.E.
Division of Basic Sciences
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N., Seattle, WA 98109-1024, USA
Tel: 206 667 4417
Fax: 206 667 6523

```



Email: pneiman@herc.org

Seq primer: T7.

## FEATURES

source

## Location/Qualifiers

1..612  
 /organism="Gallus gallus"  
 /db\_xref="taxon:9031"  
 /clone="1H9"  
 /clone\_lib="DT40subNB"  
 /tissue\_type="bursa of Fabricius"  
 /cell\_type="Avian leukosis virus-induced lymphoma"  
 /cell\_line="DT-40"  
 /dev\_stage="bursal stem cell"  
 /note="A cDNA library was prepared using both oligo dt and random primers and cloned into the Eco RI site of lambda ZAP (stratagene). This base library was subtracted with mRNA from normal 2-week bursa, and with an excess of avian leukosis virus, c-myc, and immunoglobulin sequences, by the method of Sive and St. John (A simple subtractive hybridization technique employing photoactivatable biotin and phenol extraction. Nucleic Acids Research 16, 10937, 1988) to produce DT40subNB."  
 192 a 136 c 144 g 136 t 4 others

BASE COUNT  
ORIGIN

Query Match 1.9%; Score 21; DB 24; Length 612;

Best Local Similarity 100.0%; Pred. No. 4.9; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 gaatggcaatctctcctgcc 518

Db 113 GAAGTGGCAATCTCTCTGCC 133

RESULT 8

AI599832

LOCUS

DEFINITION

EST251535 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone

REMEM21 3' end, mRNA sequence.

ACCESSION

AI599832

VERSION

AI599832.1 GI:4608880

KEYWORDS

EST.

SOURCE

Rattus sp.

ORGANISM

Rattus sp.

REFERENCE

1 (bases 1 to 495)

AUTHORS

Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,

Kerlavage,A.R. and Adams,M.D.

TITLE

Rat Genome Project: Generation of a Rat EST (RESE) Catalog &amp; Rat

Gene Index

JOURNAL

Unpublished (1998)

COMMENT

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3329

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13-21.

FEATURES

source

1..495

/organism="Rattus sp."

/db\_xref="taxon:10118"

/clone="REMEM21"

/clone\_lib="Normalized rat embryo, Bento Soares"

/dev\_stage="embryo 8, 12, 18 dpc"

/note="Vector: pT73Pac; Site\_1: EcoRI; Site\_2: NotI"

125 a 112 c 101 g 157 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

20; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY

874 ggatcaaaccttaggtctca 893

Db

21 GGATCAACCTTAGGTCTCA 40

RESULT

A2228599

LOCUS

DEFINITION

A2228599

ACCESSION

A2228599

VERSION

A2228599.1 GI:8536648

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 594)

AUTHORS

Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret

and and Fraser,C.M.

TITLE

Mouse BAC End Sequences from Library RPCI-23

JOURNAL

Unpublished (1999)

COMMENT

Other\_SSSs: RPCI-23-52L20.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301-838 0200

Fax: 301-838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html

Plate: 52 row: L column: 20

Seq primer: T7

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..594

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-52L20"

/clone\_lib="RPCI-23"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:

EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methyase. Size

selected DNA was cloned into the pBACe3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies).

187 a 126 c 99 g 182 t

BASE COUNT

ORIGIN

Query Match 1.8%; Score 20; DB 117; Length 594;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 atattgggtgcaggagctgt 368

Db 289 ATATGGGTGCAGGAGCTGT 308

```

RESULT 10
AZ028284
LOCUS      624 bp      DNA      25-FEB-2000
DEFINITION RPCI-23-312G4.TJ RPCI-23 Mus musculus genomic clone RPCI-23-312G4,
DNA sequence.
ACCESSION  AZ028284
VERSION     AZ028284.1  GI:7103668
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 624)
AUTHORS    Zhao,S., Niemman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and and Fraser,C.M.
            Email: szhao@tigr.org
TITLE      Mouse BAC End Sequences from Library RPCI-23
JOURNAL    Unpublished (1999)
COMMENT    Other GSSs: RPCI-23-312G4.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
            Plate: 312 row: G column: 4
            Seq primer: SP6
            Class: BAC ends.
FEATURES   Location/Qualifiers
            source          1..624
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-312G4"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site.1:
            EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI Methyase. Size
            selected DNA was cloned into the pBACE3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 196 a 121 c 96 g 211 t
ORIGIN
Query Match      1.8%; Score 20; DB 114; Length 624;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 atttgccttctgtgtcct 677
|||||
DB 467 ATTTGCTTTGTGTGCTT 486

RESULT 11
AZ229370
LOCUS      654 bp      DNA      GSS      14-JUN-2000
DEFINITION RPCI-23-52M21.TVB RPCI-23 Mus musculus genomic clone RPCI-23-52M21,
DNA sequence.
ACCESSION  AZ229370
VERSION     AZ229370.1  GI:8537419
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 654)
AUTHORS    Zhao,S., Niemman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and and Fraser,C.M.
            Email: szhao@tigr.org
TITLE      Mouse BAC End Sequences from Library RPCI-23
JOURNAL    Unpublished (1999)
COMMENT    Other GSSs: RPCI-23-312G4.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
            Plate: 52 row: M column: 21
            Seq primer: N7
            Class: BAC ends.
FEATURES   Location/Qualifiers
            source          1..654
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-52M21"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site.1:
            EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI Methyase. Size
            selected DNA was cloned into the pBACE3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 205 a 136 c 113 g 200 t
ORIGIN
Query Match      1.8%; Score 20; DB 117; Length 654;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 atattggttcgaggagctgt 368
|||||
DB 301 ATATGGGTGCAGGGAGCTGT 320

RESULT 12
AZ080850
LOCUS      812 bp      mRNA      EST      12-JUL-2000
DEFINITION AU080850. Sugano mouse brain mmcb Mus musculus cDNA clone MNCB-6148
5', mRNA, sequence.
ACCESSION  AU080850
VERSION     AU080850.1  GI:6085604
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 812)
AUTHORS    Hashimoto,K., Kusuda,J., Tanuma,R., Ito,A., Hirata,M., Toyoda,A.,
            Suzuki,Y., Sasaki,M. and Sugano,S.
            Isolation of full-length cDNA clones from a mouse brain cDNA
            library made by oligo-capping method
            Unpublished (1999)
            Contact: Katsuyuki Hashimoto

```

Division of Genetic Resources  
National Institute of Infectious Diseases  
23-1 Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
Email: khashienh.go.jp  
URL: http://www.nih.go.jp/yoken/genbank/

#### FEATURES

Location/Qualifiers  
1. 812  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone\_lib="WNCB-6148"  
/clone\_lib="Sugano mouse brain mncb"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="TOP10"  
/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA was primed with an oligo(dT) primer ATGCGCCCTTTTCTTTTCTTTT; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5 kb. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing: 5' end primer [CTTCTGCTCTAAAGCTGG], 3' end primer [CGACCTGCAGCTCGACACA]."

BASE COUNT 184 a 242 c 216 g 154 t 16 others  
ORIGIN

Query Match 1.8%; Score 20; DB 15; Length 812;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1061 ttatataccctacacag 1080  
DB 394 TTATATACACCTAGCACAG 413

#### RESULT 13

CNS00EFO/C 1101 bp DNA GSS 04-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #  
DEFINITION BAC29F02 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL069133  
VERSION AL069133.1 GI:4949276  
KEYWORDS GSS  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.

#### REFERENCE

AUTHORS Direct Submission  
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr)  
JOURNAL - Web : www.genoscope.cns.fr  
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org the BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Piter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library, or and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

#### FEATURES

Location/Qualifiers  
1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR29F02"  
/note="end : 77"

BASE COUNT 376 a 222 c 182 g 258 t 63 others  
ORIGIN

Query Match 1.8%; Score 20; DB 121; Length 1101;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 928 tccattctctttttttt 947  
DB 510 TCCATTCTATTCTTTT 491

#### RESULT 14

BE344801 234 bp mRNA EST 17-JUL-2000  
LOCUS 946029B/C2.Y1 946 - tassel primordium prepared by Schmidt lab Zea  
DEFINITION mays cDNA, mRNA sequence.  
ACCESSION BE344801  
VERSION BE344801.1 GI:9254333  
KEYWORDS EST.  
SOURCE Zea mays.

ORGANISM Zea mays  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.  
1 (bases 1 to 234)  
Walbot, V.  
Maize ESTs from various cDNA libraries sequenced at Stanford University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 546029 row: B column: 02.  
Location/Qualifiers  
1. 234  
/organism="Zea mays"  
/cultivar="OH43"  
/db\_xref="taxon:4577"  
/clone\_lib="946 - tassel primordium prepared by Schmidt lab"  
/tissue\_type="tassels"  
/dev\_stage="just after the transition from vegetative to inflorescence development"  
/lab\_host="XL0LR"

#### REFERENCE

AUTHORS  
TITLE  
JOURNAL  
COMMENT

#### FEATURES

Location/Qualifiers  
1. 234  
/organism="Zea mays"  
/cultivar="OH43"  
/db\_xref="taxon:4577"  
/clone\_lib="946 - tassel primordium prepared by Schmidt lab"  
/tissue\_type="tassels"  
/dev\_stage="just after the transition from vegetative to inflorescence development"  
/lab\_host="XL0LR"  
/note="Organ: tassels; Vector: HybrizAP; Site: 1: EcoRI; Site: 2: XhoI; George Chuck dissected immature tassels between 1mm and 3mm. Sharon Stanfield prepared the cDNA library in HybrizAP. Sample insert size range was 350 bp to 3 Kb with a 1 Kb average."  
BASE COUNT 36 a 75 c 81 g 75 t 42 t  
ORIGIN

Query Match 1.7%; Score 19; DB 35; Length 234;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 agaaggccctgcctact 208  
|||||

DB 215 AGAGGGCGCTGCTACT 233

RESULT 15  
 AQ641537/C  
 LOCUS  
 DEFINITION  
 AQ641537 321 bp DNA GSS 08-JUL-1999  
 RPCI93-DpnII-26D10.TJ RPCI93-DpnII Trypanosoma brucei genomic clone  
 RPCI93-DpnII-26D10, DNA sequence.  
 ACCESSION  
 AQ641537  
 VERSION  
 AQ641537.1 GI:5118247  
 KEYWORDS  
 GSS.  
 SOURCE  
 Trypanosoma brucei.  
 ORGANISM  
 Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.  
 REFERENCE  
 1 (bases 1 to 321)  
 El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,  
 Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,  
 Fraser, C. and Adams, M.  
 Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93  
 Library for gene discovery and sequence-ready map construction  
 Unpublished (1999)  
 Other GSSs: RPCI93-DpnII-26D10.TV  
 Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org  
 Clones and high density filters may be purchased from BACPAC  
 Resources (<http://bacpac.med.buffalo.edu>). BAC end sequences search  
 page: <http://www.tigr.org/tdb/mbd/tbdb/>.  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 Source  
 1..321  
 Location/Qualifiers  
 /organism="Trypanosoma brucei"  
 /strain="TREU927/4 GUTat 10.1"  
 /db\_xref="taxon:5691"  
 /clone="RPCI93-DpnII-26D10"  
 /clone\_lib="RPCI93-DpnII"  
 /note="Vector: pBACE3.6; Site\_1: Bam HI; Site\_2: Bam HI;  
 Constructed for the Institute for Genomic Research by  
 Bonhui Zhao in Pieter de Jong's laboratory (Roswell Park  
 Cancer Institute, Buffalo, NY). Briefly, Trypanosoma  
 brucei TREU927/4 GUTat 10.1 agarose embedded DNA was  
 partially digested with a combination of Eco RI and Eco RI  
 methylase (RPCI93-EcoRI segment) or Dpn II (RPCI93-DpnII  
 segment). High molecular weight fragments were ligated in  
 pBACE3.6 vector digested with Eco RI or Bam HI,  
 respectively. The average insert size is 141 Kb. Total  
 coverage (both segments): > 90 X the haploid  
 non-minichromosomal genome."  
 BASE COUNT 98 a 82 c 80 g 61 t  
 ORIGIN

Query Match 1.7%; Score 19; DB 100; Length 321;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 930 cattctattttctttgt 948  
 |||||||  
 Db 103 CATTTCATTATTTCTTTGT 85

Search completed: November 4, 2000, 01:08:59  
 Job time: 17440 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 01:24:20 ; Search time 7109.34 Seconds  
(without alignments)  
669.635 Million cell updates/sec

Title: US-09-101-423a-6  
Perfect score: 1090  
Sequence: 1 ttgtctctgtgttactgtt.....cctagcacagtcacccaca 1090

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_bai.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pli.\*  
8: gb\_pl2.\*  
9: gb\_pr1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: em\_fun.\*  
13: em\_hum1.\*  
14: em\_hum2.\*  
15: em\_in.\*  
16: em\_om.\*  
17: em\_or.\*  
18: em\_ov.\*  
19: em\_pat.\*  
20: em\_ph.\*  
21: em\_pli.\*  
22: em\_pl2.\*  
23: em\_ro.\*  
24: em\_sts.\*  
25: em\_sy.\*  
26: em\_un.\*  
27: em\_vi.\*  
28: gb\_hgt1.\*  
29: gb\_hgt2.\*  
30: gb\_in2.\*  
31: em\_bai.\*  
32: em\_ba2.\*  
33: em\_hum3.\*  
34: em\_hum4.\*  
35: gb\_pr4.\*  
36: gb\_hgt3.\*  
37: gb\_hgt4.\*  
38: gb\_hgt5.\*  
39: gb\_hgt6.\*  
40: gb\_hgt7.\*  
41: em\_hgt1.\*  
42: em\_hgt2.\*  
43: em\_hgt3.\*

44: em\_hum5.\*  
45: gb\_pl3.\*  
46: gb\_pr5.\*  
47: gb\_hgt8.\*  
48: gb\_hgt9.\*  
49: gb\_hgt10.\*  
50: gb\_hgt11.\*  
51: gb\_hgt12.\*  
52: gb\_hgt13.\*  
53: gb\_hgt14.\*  
54: gb\_in3.\*  
55: gb\_hgt15.\*  
56: gb\_hgt16.\*  
57: gb\_hgt17.\*  
58: em\_hgt4.\*  
59: em\_hgt5.\*  
60: em\_hgt6.\*  
61: em\_hgt7.\*  
62: em\_hum6.\*  
63: gb\_hgt18.\*  
64: gb\_hgt19.\*  
65: gb\_ba3.\*  
66: em\_hgt8.\*  
67: em\_hgt9.\*  
68: em\_hgt10.\*  
69: em\_hgt11.\*  
70: em\_hgt12.\*  
71: em\_hgt13.\*  
72: em\_hgt14.\*  
73: em\_hgt15.\*  
74: em\_hgt16.\*  
75: em\_hgt17.\*  
76: em\_hgt18.\*  
77: em\_hgt19.\*  
78: em\_hgt20.\*  
79: em\_hgt21.\*  
80: em\_hgt22.\*  
81: em\_hgt23.\*  
82: gb\_pr8.\*  
83: gb\_pr7.\*  
84: gb\_hgt20.\*  
85: gb\_hgt21.\*  
86: gb\_hgt22.\*  
87: gb\_hgt23.\*  
88: gb\_ro.\*  
89: gb\_sts1.\*  
90: gb\_sts2.\*  
91: gb\_sy.\*  
92: gb\_un.\*  
93: gb\_vil.\*  
94: gb\_vil2.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1090	100.0	1090	5	A64063	A64063 Sequence 6
2	260.8	23.9	13222	88	RATCP4A2	M57719 Rat cytochr
3	226.2	20.8	16257	88	RNVAOX	X59496 Rat genes f
c 4	215.6	19.9	1305	88	AF000199	AF000199 Rattus sp
c 5	213.8	19.6	1429	88	RN8MLV1	U87598 Rattus norv
c 6	211.6	19.4	1857	88	RNLB6	X07886 Rat L1Rn B6
c 7	205.8	19.0	2288	88	RNLB7	X07887 Rat L1Rn B7
c 8	202.8	18.6	1986	88	RN14MLV11	U87602 Rattus norv
c 9	202	18.5	1297	88	RN11MLV11	U87600 Rattus norv
c 10	184.2	16.9	1297	88	AF000204	AF000204 Rattus sp
c 11	172	15.8	6335	88	RNLNED	X53581 R. norvegicu
c 12	155.8	14.3	1117	88	AF000202	AF000202 Rattus sp

c 13 150.4 13.8 1127 88 AF000203  
 c 14 149.6 13.7 1127 88 AF000200  
 c 15 147.6 13.5 1113 88 RN38N1V11  
 c 16 142.8 13.1 1821 88 RN56N1V11  
 c 17 141.8 13.0 1102 88 AF000205  
 c 18 132.2 12.1 5577 88 RNRL13  
 c 19 128.4 11.8 1826 88 RNRL1P5F  
 c 20 121.8 11.2 2225 88 AF058787  
 c 21 121.8 11.2 3047 88 RATEFLAA  
 c 22 121.8 11.2 8048 88 RATLIN3A  
 c 23 97 8.9 2946 88 S51970  
 c 24 93.2 8.6 7072 88 RATPRSTTNC  
 c 25 91.6 8.4 948 88 AF000201  
 c 26 85.4 7.8 6369 88 RATCRYGF  
 c 27 80.2 7.4 397 89 AU046833  
 c 28 79.4 7.3 433 89 AU026950  
 c 29 76.8 7.0 2711 88 AB017578S2  
 c 30 74.6 6.8 344 89 AU047817  
 c 31 72.8 6.7 3635 88 AF087675  
 c 32 68.2 6.3 532 89 AU027295  
 c 33 67.8 6.2 3030 88 S53987  
 c 34 65.4 6.0 454 88 S77423  
 c 35 63.8 5.9 384 89 AU024858  
 c 36 60.6 5.6 406 89 AU047071  
 c 37 55.6 5.1 756 89 AF052311  
 c 38 52.6 4.8 206117 47 AC021063  
 c 39 49.8 4.6 210861 57 AC073734  
 c 40 48.8 4.5 163639 88 AC020974  
 c 41 48.2 4.4 365 89 AU048612  
 c 42 46.2 4.2 289065 52 AC027298  
 c 43 45 4.1 224142 50 AC025587  
 c 44 44.6 4.1 71564 52 AC034284  
 c 45 44.6 4.1 71564 52 AC034284

## ALIGNMENTS

RESULT 1  
 A64063  
 LOCUS A64063 1090 bp DNA  
 DEFINITION Sequence 6 from Patent WO9725443.  
 ACCESSION A64063  
 VERSION A64063.1 GI:3717537  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCES  
 1 (bases 1 to 1090)  
 Rudland, P.S. and Barraclough, B.R.  
 TITLE METASTASIS INDUCING DNA S  
 JOURNAL Patent: WO 9725443-A 17-JUL-1997;  
 UNIV LIVERPOOL (GB)  
 FEATURES  
 source  
 1..1090  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 224 a 277 c 275 g 314 t  
 ORIGIN

Query Match 100.0%; Score 1090; DB 5; Length 1090;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1090; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttgtctctgtgttactgttttccattctgtacagtggttgaccttctatacgcctg 60  
 Db 1 TTGTCTCTGTGTGTTACTGTTTCCCTTCTGACAGTGGTTGACCTTCTATACGCTG 60  
 Qy 61 tgtgtcaggagtgctagacactattctctgtttttttccagcagttacaggaacaga 120  
 Db 61 TGTGTCTGAGGAGTGTGTAGACCTATTTTCTGTGTTTCTTTTTCAGCCAGTTACAGGAACAGA 120

Qy 121 gtgttctactjtcagatgtgttagctgttctctgttcactgacttctcaagctgtctctgtgt 180  
 Db 121 GTGTCTTACTTCTCAGATGTGTAGTGTCTCTGTCTCTGTCTCTCAAGCTGTCTGTGT 180  
 Qy 181 gcaggaaacaaagggcgtctccctacttctacttctggccctacgcacagaggggctcaga 240  
 Db 181 GCAGGAACCAAGAGGGCGCTGTCCCTACTTCTACTGGCCCCCTACGCACAGGGGGCTAGA 240  
 Qy 241 tgggtctagtggttttctctctagagcctgaatgtgggcagagtagtctctctcgtgtt 300  
 Db 241 TGGTCTTAGGTGTTTCTCTCTAGAGCCTGAATGTGGCAGACAGTAGTCTCTCTCTGGT 300  
 Qy 301 tctctagtagtcttct 360  
 Db 301 TCCTAGTAGTATGTCT 360  
 Qy 361 ggaagctgttct 420  
 Db 361 GGAGCTGTTTACCAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420  
 Qy 421 ctgtctctgtctcaatct 480  
 Db 421 CTTGCTCTCTCTCAATCT 480  
 Qy 481 gtggcaaaagtgtggcagaagtgtggaatctctctctctctctctctctctctctctct 540  
 Db 481 GTGGCAAAAGTGGGAGAGTGGCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540  
 Qy 541 actctgggggaatccgct 600  
 Db 541 ACTTCTGGGCAATCCGCT 600  
 Qy 601 atcaggcaaaagtgttgagcaaacagtaggaactggaagtgtcaggtctccagaggaatt 660  
 Db 601 ATCAGCAAAAGTGTGAGCAAACTAGAACTGGAAGTGTGAGTGTGAGTGTGAGTGTG 660  
 Qy 661 ttgcctctgtgtctctgagtcaccagcagctgacttggagcagagaaatgtgttttc 720  
 Db 661 TTGCTTGTGTGTCTGTAGTCTCAGTCCAGGAGTCTCTCTCTCTCTCTCTCTCTCTCT 720  
 Qy 721 cctcgtctcagggctgaagtgtgacacctcaggggttggtcttccagctgtacctgtggaaa 780  
 Db 721 CCTCGGTCTCAGGCTGAAAGTGTGACCTCAGGCTTGGCTTTCAGCTGTACCTGTGGA 780  
 Qy 781 gtatgttttttaaaatcttaagtagctatctacgcagcagaggtgtgttaaatgtctatt 840  
 Db 781 GTATGTTTTTAAATCTAAGATAGTATCATATGACAGAGGCTTGTGTAATATGCTATT 840  
 Qy 841 tgggtctcttatgacttactttgtgttactgagatcaaacctaggtgtctcaagcagtc 900  
 Db 841 TGGTCTCTTTATGACTTACTTTTGTGTACTGATCCAGCTCCATTTCTATTCTTTTGT 900  
 Qy 901 atcaaatctctgtcactgtatccagctcattctctctctctctctctctctctctctct 960  
 Db 901 ATCACAATTTCTGTCTACTGTATCCAGCTCCATTTCTATTCTTTTGTCTCCGCG 960  
 Qy 961 ctgcgcagcaagaaacacgcgtacgacatcagcaatcctctctctctctctctctctct 1020  
 Db 961 CTCGCCAGCAAGAAACACGCTAGGACATACGAATCTCTCTCTCTCTCTCTCTCTCT 1020  
 Qy 1021 tgaattcttaagagagcccgccagcagcagctgtggcgcggtttatatcacacctacacag 1080  
 Db 1021 TGAATCTTTAAAGAGAGCCCGCCAGCTGCGGCTTTTATATACACCTTAGCAGCAG 1080  
 Qy 1081 tgcattccaca 1090  
 Db 1081 TGCATCCACA 1090

RESULT 2  
 RATCYP4A2  
 LOCUS  
 DEFINITION Rat cytochrome P-450 IVA2 (CYP4A2) gene, complete cds. 27-APR-1993

[illegible]





```

Db 10750 ACTGTTATCCGGTCTGTTTCCCTCAGATTCCGGGGTGTCTCAGGCAGGGGTCCTCAGC 10809
Qy 422 ttgctgtgcaaatcttccgcaccacagagc--acccaagtctctctcttgggccaagga 479
Db 10810 CTCCTGGCCCTCCGCCAGCAACCCAGAGGCTTATACAGTTTCTCTTGGGCCAGGGA 10869
Qy 480 tctgggcaaaagtgggcaagaagtggcaatctctctcccttagctctcagattgacct 539
Db 10870 TGTGGCAGGGGTGGCAGTGTGGTGTCTCTCCGCTCTGCAGCCTCAGGAGTGCCCA 10929
Qy 540 caactctgggcaatccgctctctctccacaggggtttgggagcagggagctgtgggcgg 599
Db 10930 CCTGACAGGGCGGTGAGTCTCTCTCAGCGGTCTGGGAGCAGAGAGCTGCTG--CGG 10987
Qy 600 tatcaggcaaaagtgttgaggcaacaggttagaaactggaagtgcaggtccacagagaat 659
Db 10988 GCGGGATCCCGGGGTGTGGACCTCGGCAACACACAGGACGTCGCCGCTCTAGATGAAC 11047
Qy 660 ttgctctgtgtctcctgagtcacacagcagcaggtcacttgagcagaaaaattggttt 719
Db 11048 TCTGCTGTGTGTGCCAATCTCACCAGGCGCTCTCTTCGACGAGCAAGTTGGTCTT 11107
Qy 720 cccctggtctcagcctgaattg 744
Db 11108 ACTGTGTCGGAGGCTCAAGTTG 11132

RESULT 4
AF000199/c
LOCUS
DEFINITION
Rattus sp. T-612.
Rattus sp. T-612.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1306)
Cabot,E.L., Angeletti,B., Usdin,K. and Furano,A.V.
Rapid evolution of a young L1 (LINE-1) clade in recently speciated
Rattus taxa
J. Mol. Evol. 45 (4), 412-423 (1997)
97472947
MEDLINE
REFERENCE
2 (bases 1 to 1306)
Cabot,E.L. and Furano,A.V.
Direct Submission
TITLE
Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA
FEATURES
source
1. .1306
/organism="Rattus sp. T-612"
/isolate="T-612"
/db_xref="taxon:68896"
/notes="from the collection of Institut des Sciences de
l'Evolution, Montpellier II; collected from the island of
Ceram, Indonesia and identified by Djoko Iskandar as
Rattus moluccarius; cranial morphology suggests similarity
to Rattus rattus satarae"
1. .921
5'UTR
repeat_region 1. .>1306
/notes="u-rm2"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type="dispersed
922. .>1306
CDS
/notes="orf 1; putative RNA binding protein"
/codon_start=1
/pseudo
BASE COUNT 441 a 351 c 311 g 199 t 4 others
ORIGIN

```

```

Query Match 19.9%; Score 216.6; DB 88; Length 1306;
Best Local Similarity 65.5%; Pred. No. 2.7e-57;
Matches 391; Conservative 0; Mismatches 198; Indels 8; Gaps 5;

Qy 165 caagctgtctctgtgtgaggaacacagagggccctctactctactctg-ggcccta 223
Db 639 CGAGCTCTCTGCGGGGCGACCAACAGGAGATCTGTGCCGCCCTTCGGGAGCTCTG 580
Qy 224 cgcacaggggggctagatggtgctaggtgttttctctctagagcctgaaatgtgggcagag 283
Db 579 TGCACATAGGCTTCCAGAGGCGCTTGGTGTCTCTCTGCGCTCCGAGATGTGTGCAGAG 520
Qy 284 agtagctctctggttttcttaggtatgtctccctctgaaaggtctagctctccctcc 343
Db 519 TGCAGTCTCTCTGCTGTTTCCCA-GCATGCTGCTCTCTGAAAGTTAGTTCTCCCTCCC 461
Qy 344 atgggatagggtgcagggagctgtt----tgaccaggtctctcaaaatcccggtgcagt 399
Db 460 ACAGGATTGGGGCGCAGAGATTTAGTGTGCTGTCTCTTCCATCAGGTTCCGGGGTGT 401
Qy 400 ctggacgca-ggctccttagctgtcgtgcaaatctcccgacccagagggaccca 458
Db 400 CTCAGGCGCAGGATCTCTGCGCTCTGCGGCCCTCTACGGGAACCCAGAGGCTTATAC 341
Qy 459 agttctcttgggccaaggtatgtggcagaagtggtggcagaagtgcaatctctctctcc 518
Db 340 AGTTCTCTTGGGCCAGAGATGTGGCAGGGGTGGGCAGTGTGTGGTCTCTCTCCGCT 281
Qy 519 ctgagctctaggttgccctcactctctgggcaatcgtctctctctccacaggggttgg 578
Db 280 CTCAGGCTCAGGAGTGCCACCTCACCAGGCGGTGAGTCTCTCTCCATGGGGTCTGG 221
Qy 579 gagcaggaagctgtggcggtatcagggcaaaagtttgaggaacacagtagaaactga 638
Db 220 GACCATTAGCTCTGCGGGCGGATTCGCGGGTTG-GGACTCTGTGTAACACAGGA 162
Qy 639 agtgcaggtcccagaggaatttgcctttgtgtcctgagtcctccacagggaggtcact 698
Db 161 ACTNNAGGTCCTAGAGGAATCTNCTCTGTCTATCCGAGTTCACAGGAGGCTCTCT 102
Qy 699 tggagcagaaaaattggttttccctcctggtctcagcctgaaagtgcacctcaggt 755
Db 101 TGCAGCAGAAAGTTGTCTTACTGTGTGTTCGGAGGCTCAAGTTTCNCGCGGGGT 45

RESULT 5
RN8MLV1/c
LOCUS
DEFINITION
Rattus norvegicus L1 retrotransposon mlv12-rn8, 5'UTR and putative
RNA binding protein 1 gene, partial cds.
ACCESSION
U87598
VERSION
U87598.1 GI:2039337
KEYWORDS
1 of 2
SEGMENT
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1429)
Hayward,B.E., Zavanelli,M. and Furano,A.V.
Recombination creates novel L1 (LINE 1) elements in Rattus
norvegicus
Genetics (1997) In press
REFERENCE
2 (bases 1 to 1429)
Hayward,B.E., Good,B., Cabot,E. and Furano,A.V.
Direct Submission
TITLE
Submitted (29-JAN-1997) LMCB/NIDDK, NIH, 8/203 8 Center Dr. MSC
0830, Bethesda, MD 20892-0830, USA
FEATURES
Location/Qualifiers
1. .1429
/organism="Rattus norvegicus"

```

```
/db_xref="taxon:10116"
1. .>1429
/loc="mlv12-rn8"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type="dispersed"
1. .>910
911. .>1429
/note="ORF I"
/codon_start=1
/product="putative RNA binding protein 1"
/protein_id="AAB2995.1"
/db_xref="GI:2039340"
/translation="MARGKRNPNSRNQDYMASSEPNSTKTNMEYNTPEKQDLVSK
SYLIMLDFDKVNSLREQVEDYREESQRLKEFOENIKQVEAHRESQKSLREF
QKSLKEYEYLNKQVEAHRESQKSLKEFOENTIKQLKELMEIAIKKHEMETTLDI
ENOKKRGAVDT"
BASE COUNT 505 a 367 c 349 g 208 t
ORIGIN
Query Match 19.6%; Score 213.8; DB 88; Length 1429;
Best Local Similarity 66.0%; Pred. No. 2.1e-56;
Matches 389; Conservative 0; Mismatches 187; Indels 13; Gaps 5;
QY 165 caagctgtctctgtgacgaacacagaagggcctgtccctacttctac-tggggcccta 223
DB 631 CAGGCTCTCTGACGGCGGACGAGCAACGAGACACTCTGCCGCCCTTCGGGGGCTTCAG 572
QY 224 cgcacaggggctagatggtgctaggtgtttctctctagacctgaatg--tgggcag 281
DB 571 TGCACAGGGTTCAGATGGCTTTGGGCTTTTCTCTCTGGCGTCCGAGATCATGTCAG 512
QY 282 agagtagtctctctgttctctagtagtcttctccctctgaggtctagctctccct 341
DB 511 AGACAGCTCTCTCTGTTTCCAGGCTGTGCTGCCCTCTGAGAGTTAGCTCTCCCTC 452
QY 342 ccaatggataggtgacaggaagctgtttaccagggtctctcaaatccgggtgcagtct 401
DB 451 CCACGGGATTTGGGTGACAGAACTGTTTATCCAG--TCTGTTTTTTTTCAGGTTCCGGCG 394
QY 402 ggaacagaggtctctgtagctgtctgctgctgcaatctccc-----gcacccagagcca 454
DB 393 GTGCTCAGGACAGGGGCTCTGCCGCTCTGGGCGCTCCGCCACGGGAGCCCGAGGCGCTT 334
QY 455 cccaagtctctctgtggcccaagatggtgggcaaaagtggtggcagaagtggcaatctctcc 514
DB 333 ATACAGTTTCTCTTGGGCCAGGATGTGGCGAGGGGTGGGCACTGTGTGGTCTCTTC 274
QY 515 tgcctagctctcagatggtcctcactctctgggcaaatccgctctctctctccacaggggt 574
DB 273 TGCTCTCAGGCTCAGGAGTCCCACTGACCGAGGGCGGTTGGGTCTCTCTCAGCGGGT 214
QY 575 ttgggagcagggagctgtgggcccgtatcagggcaaaagttgtgaggaaccagctagaac 634
DB 213 CTGGGACAGAGAGCTGCTGGGGCCGGGATCCCGGGGTGGGACTTCCAG-TAAACAC 155
QY 635 tggaggtcaggtccacaggaattgtgctctgtgctctgagtcacacagggcaggt 694
DB 154 AGACAGTTCGCGGTCTAGAGGAATCTGCTTCGCTGTGTCCTCCAAAGCTACACGAGCAGCT 95
QY 695 cacttgagcagaaaaattgttttccctcgctctcagcctgaagtt 743
DB 94 TTCTTGACGAGAAAGTTGGTCTTACCTGTGTGTCGAGGCTCAAGTT 46
```

## RESULT 6

```
RNLB6/c
LOCUS RNLB6 1857 bp DNA ROD 06-JUL-1989
DEFINITION Rat L1Rn B6 repetitive DNA element.
ACCESSION X07686
VERSION X07686.1 GI:56554
KEYWORDS L1 repetitive sequence; long interspersed repeat; repetitive
sequence; tandem repeat.
```

```
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 1857)
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
TITLE Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
JOURNAL Furano A.V.
Direct Submission
Submitted (16-MAY-1988) Furand A.V., NIH, 81203, Bethesda, MD
20892, USA
REFERENCE 2 (bases 1 to 1857)
AUTHORS Furano A.V., Robb S.M. and Robb F.T.
TITLE The structure of the regulatory region of the rat L1 (L1Rn, long
interspersed repeated) DNA family of transposable elements
JOURNAL Nucleic Acids Res. 16 (19), 9215-9231 (1988)
MEDLINE 89016625
FEATURES
Location/Qualifiers
1. .1857
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
misc_feature 1. .461
/note="non L1 flanking DNA"
misc_feature 462. .559
/note="partial tandem duplication of L1Rn B6 DNA"
promoter 560. .1167
/note="L1Rn B6 DNA promoter"
CDS 1358. .>1857
/note="ORF1 (1857 is 2nd base in codon)"
/codon_start=1
/protein_id="CAA30528.1"
/db_xref="GI:56555"
/db_xref="SPTREMBL:O63297"
/translation="MARGKRNPNSRNQDYMASSEPNSTKTNMEYNTPEKQDLVSK
SYLIMLDFDKVNSLREQVEDYREESQRLKEFOENIKQVEAHRESQKSLREF
QKSLKEYEYLNKQVEAHRESQKSLKEFOENTIKQLKELMEIAIKKHEMETTLDIENOKKRG
GAVDT"
repeat_region 1548. .1679
/note="direct repeat"
repeat_region 1680. .1745
/note="direct repeat"
BASE COUNT 623 a 462 c 418 g 354 t
ORIGIN
Query Match 19.4%; Score 211.6; DB 88; Length 1857;
Best Local Similarity 65.1%; Pred. No. 1.1e-55;
Matches 392; Conservative 0; Mismatches 199; Indels 11; Gaps 5;
QY 164 tcaagctgtctgtgtgacgaacacagaagggcctgtccctacttctacttg-gccct 222
DB 1076 TCGAGCTCTCTCGGCTGCAGCAACGAGAGATCTGGCGCGCTGTTTCCAGGAACITCA 1017
QY 223 acgcacaggggcttagatggtgctaggtgtttctctctagacctgaatctgtg--gca 280
DB 1016 GTGACACAGGGTTCAGATGGCGGTGTGTGTCTTCTCTGCGATCCGAGATGTGTGCA 957
QY 281 gagagtagtccctctgttctctaggtatgtctccctctgaagctctagctctccct 340
DB 956 GACTCAGCTCTCTGCGTTCAGGATGTCTGCGCTCTCTGAGGTTTACGCTCTCTCT 897
QY 341 tccatggatattgggtgagggagctgtttgaccag-----gtcctctcaaatccgggtg 395
DB 896 CCCACGGGATTTGGGTGCAGAGAACTGTATTATCGGGTCTGTCTTCTCAGGTTCCGGCGG 837
QY 396 cagcttgac-gcaggctcctctagctgctgcaatcttcccgcacccacagggc-- 453
DB 836 TGCTCAGAGAAAGAGGTACTGCCACTCTCTGGGCCCTCCCGGAAACGAGAGGCT 777
QY 454 acccaagtctctcttgggccaagatgtgggcaaggtgggcaagtgaggaagtggaatctctc 513
DB 776 TATACAGTTTCTCTCTGGGCCAGGGATGTAGGAGGGGTGGGAGTGTGGTGTCTATT 717
QY 514 ctgcccctagctctcagagattgcccctcactctctgtgggaatccgctctctcttccacaggg 573
```

```

Db 716 CCGCTCTGCAGCCTCAGGAGTCCACCTGACACAGGTGGTGGGTCTCTCTCCACGGTG 657
QY 574 ttgtggagcaggagctgtgtggccggtatcaggaacaaaggtttgaggaacacagtagaaa 633
Db 656 TCTGGAGCAAGAGCTGCTCGGCGCGGATCCGCGGTGTG-GGACTCCCGGTAACA 598
QY 634 ctgaaagtgtcaggtccagaggaatttgccttctgtgtcctcagtcacacacagcagg 603
Db 597 CAGGAGGTGCGCGCTCTAGAGAAATTCTGCTGTGTGTCCTCCAGGATCACCAGGCAGG 538
QY 694 tcaactgagcagaaaattgtttccctcgcgtctcagggcctgaagtgtgcacctcagg 753
Db 537 TCTCTGACAGAAAAGTTGGTCTTACCTGTGTGTCCTCCAGGCTCAAGTTGTCTACGGG 478
QY 754 gt 755
Db 477 GT 476

RESULT 7
RNLB7/c 2288 bp DNA ROD 04-APR-1990
LOCUS Rat L1rn B7 repetitive DNA element.
DEFINITION X07687
ACCESSION X07687
VERSION 1
KEYWORDS L1 repetitive sequence; long interspersed repeat; repetitive
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
TITLE 1 (bases 1 to 2288)
JOURNAL Furano, A.V.
Direct Submission
Submitted (16-MAY-1988) Furano A.V., NIH, 81203, Bethesda, MD
20892, USA
REFERENCE 2 (bases 1 to 2288)
AUTHORS Furano, A.V., Robb, S.M. and Robb, F.T.
TITLE The structure of the regulatory region of the rat L1 (L1rn), long
interspersed repeated) DNA family of transposable elements
JOURNAL Nucleic Acids Res. 16 (19), 9215-9231 (1988)
MEDLINE 89016625
COMMENT Data kindly reviewed (11-JAN-1989) by FURANO A.V.
FEATURES
source
1..2288
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..511
/misc_feature
/feature="non L1 flanking DNA"
512..1126
/misc_feature
/feature="partial tandem duplication of L1rn B7"
1127..1731
promoter
/feature="L1rn B7 DNA promoter"
1921..>2288
CDS
/feature="ORF1 (2288 is 2nd base in codon)"
/codon_start=1
/protein_id="CAA30529.1"
/db_xref="GI:56557"
/translation="MARGKRNPSNRQDYMASPEPNTKANTPEYNTPEKQDLVSK
SYLWLEDFKDKIDKNSLREQVAYRESQSLKEFOENTIKQLKELMEIEAIKKEH
METLIDENOKKROGAVDTs"
BASE COUNT 652 a 622 c 553 g 461 t
ORIGIN

Query Match 19.0%; Score 206.8; DB 88; Length 2288;
Best Local Similarity 64.1%; Pred. No. 3.7e-54;
Matches 377; Conservative 0; Mismatches 202; Indels 9; Gaps 4;

QY 175 ctgtgtgacaggaacagagagcctgtcccttacttcttacttctgtggtccctcagcagcagggg 234
Db 1628 CTGTGGCAGCAACCGGGAATCTGTGCCGCTCTCTCCAGGAGCCTCCGTGACACAGGCT 1569

```

```

QY 235 cctagatgtgtagtttttttctctctagagcctgaattgtgggcagagtagtctct 294
Db 1568 TTCAGATGACGTTTGGTGTTCCTCTGGCCCTG-GAATGTCACAGAGTGACGTCTCTT 1510
QY 295 ctgtgttctcaggtatgtcttccctctgaaggttagctctcctccctcagtggtatgg 354
Db 1509 CTGGTTTCCAGCGCTGTCGCGCTCTCTGAAGGTTTCAGCTCTCCCTCCCTCGGATTTGG 1450
QY 355 gtgcagggagctgttttgaccag-----gtcctctcaaatccgggtgcagttctggaccgca 409
Db 1449 GTGCAGAGAACTGTTTATCCGGTCTGTTCCTTCAGGTTCCGGCAGTGTCTCAGCGGCAG 1390
QY 410 ggtcctctgctgtgtcgtcgtcaattctcccgccacccagagggc--accacagtttctctc 467
Db 1389 GGGTCTGCGGTCCTCCGGCCCTCCCTACGGGACACACAGAGGCCCTTATACAGTTGCTCTC 1330
QY 468 ttgggccaaggtatgtggcacaaggtgtggcagaagtggaactctctcctcctcctagcgtct 527
Db 1329 TTGGGCAGCGATGTGGCAGGGGTGGGCGAGTGTGTGGTCTCTCCCTCTCGAGCCT 1270
QY 528 caggattgcctcaactctgtggcaatccgctctctctccacaggttttggcagcagga 587
Db 1269 CAGGAGTGCCACCTGATCAGGCGGTGGGGTCTCTCCACGCGGTTTGGAGGACAGAGA 1210
QY 588 gctgtgtggccggtatcaggcacaaggtttgaggaacacaggttagaaactggaagtgtcagg 647
Db 1209 GCTGCTGCGGTCCGGGATCCGGCGGTGTGGACTTCC-GATAAACACAGAAAGTGCCCGG 1151
QY 648 tccagaggaatttgccttctgtctcgtcgtcagtcacacagcagcagtcacttggagcaga 707
Db 1150 CCTAGAGGAATTTGCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1091
QY 708 aaatgtgtttccctcgtcgtcagcctcagctgaagtgtcagcctcaggt 755
Db 1090 GAAGTGGTCTCTACCTGACAGATCCAAAGGCTCAAGTTTGTCTGCTGGGGT 1043

RESULT 8
RNLAMVL1/c 1986 bp DNA ROD 25-APR-1997
LOCUS Rattus norvegicus L1 retrotransposon mlv12-rn14, 5'UTR and putative
DEFINITION RNA binding protein 1 gene, partial cds.
ACCESSION U87602
VERSION U87602.1
KEYWORDS GI:2039341
SEGMENT 1 of 2
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
TITLE Rattus.
JOURNAL 1 (bases 1 to 1986)
AUTHORS Hayward, B.E., Zavanelli, M. and Furano, A.V.
TITLE Recombination creates novel L1 (LINE 1) elements in Rattus
norvegicus
JOURNAL Genetics. (1997) In press
REFERENCE 2 (bases 1 to 1986)
AUTHORS Hayward, B.E., Good, B., Cabot, E. and Furano, A.V.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1997) LMCB/NIHDK, NIH, 8/203, 8 Center Dr. MSC
0830, Bethesda, MD 20892-0830, USA
FEATURES
source
1..1986
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..922
5'UTR
repeat_region
1..>1986
/feature="mlv12-rn14"
/feature="LINE 1 non-LTR retrotransposable element"
/feature="rpt_type-dispersed"
923..>1986
/codon_start=1
CDS

```



Mon Nov 6 10:14:43 2000

```

Db 92 GCAGCAGAAATTTGGTCTTACCTGTGTCCGAGGCTCAGGTTCCGCTCGTGGGGT 37
|||||
RESULT 10
AF000204/c 1297 bp DNA ROD 11-DEC-1997
LOCUS Rattus sp. T-612 retrotransposon mlv12-rml3 5'UTR and putative RNA
DEFINITION binding protein gene, partial cds.
ACCESSION AF000204
VERSION AF000204.1 GI:2599316
SOURCE Rattus sp. T-612.
ORGANISM Rattus sp. T-612.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1297)
Cabot, E.L., Angeletti, B., Usdin, K. and Furano, A.V.
Rapid evolution of a young L1 (LINE-1) clade in recently speciated
Rattus taxa
J. Mol. Evol. 45 (4), 412-423 (1997)
MEDLINE 97472947
REFERENCE 2 (bases 1 to 1297)
Cabot, E.L. and Furano, A.V.
Direct Submission
JOURNAL Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA
FEATURES
source
Location/Qualifiers
1..1297
/organism="Rattus sp. T-612"
/isolates="T-612"
/db_xref="taxon:68996"
/note="from the collection of Institut des Sciences de
l'Evolution, Montpellier II; collected from the island of
Ceram, Indonesia and identified by Djoko Iskandar as
Rattus moluccarius; cranial morphology suggests similarity
to Rattus rattus sataeae"
repeat_region
1..>1297
/note="mlv12-rml3"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type="dispersed"
1..913
914..>1297
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein"
/protein_id="AAC53499.1"
/db_xref="GI:2599317"
/translation="MARGKRNPSNRNODMPSPNPTKTNMEYPTPEKQDLVSK
SYLIMLEDFKKDNNILRETFQELINKQVEAYREWQSKLKEFQENTIKMKELKMEIE
AIKKHEMETLIDENKRRQGVDTs"
BASE COUNT 434 a 356 c 324 g 183 t
ORIGIN
Query Match 16.9%; Score 184.2; DB 88; Length 1297;
Best Local Similarity 64.5%; Pred. No. 5.1e-47;
Matches 389; Conservative 0; Mismatches 198; Indels 16; Gaps 7;
QY 168 gctgtctctgtgcaggaaccagagggcctgtctctctacttctactgg-gccctacgc 226
|||
Db 631 GCTCTCTGAGCGGCGAGCAACGAGACCTGTGCCGCCCTTCCGGGAGCTTCAGTGC 572
|||
QY 227 acagggggcctagatggctgctagggtgtttctctctagacctgaatgtgg-gcagaga 284
|||
Db 571 ACCAGGGTTCAGATGGTCTTTGGCTTTTTCCTCTGGCGTCCGAGATGTGTGTCAGGGA 512
|||
QY 285 gtatgtctctctgggtt--cctaggatgtctctctctctctctctctctctctctc 342
|||
Db 511 GCAGTCTCTCTCTGTTTCCCAAGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 452
|||
QY 343 catgggatatgggtgcagggagagctgtttgaccagaggtctctctc-----aaatcggggtg 395
|||||
Db 451 CACGGGATTTGGTGCAGAGACTGTTTATCCGGTCTGTTTCTTTTTCAGGTTCTGGCGCG 392
|||||
QY 396 cagctggaccgcag-gctcctgttagcttgcctgtgcaaatcttcccgaccagagggc- 453
|||||
Db 391 GTGTCTCAGGCGCAGAGGTCCTGCCGCTCTCGGCCCTCCGCCACGGGAGCCAGAGGCC 332
|||||
QY 454 -accacaatttccctctgggccaagagatgtggccaaggtgggccaagtggaatctct 512
|||||
Db 331 TTATACAGTTTCTTCTTGGGCCAGGATGTGGTAGGGGTGAGCAGTGTGTGGTCTCT 272
|||||
QY 513 cctgcctcagctgtcagagattgccctcacttctgtggcaatccgctctctctctccacagg 572
|||||
Db 271 TCGGCTCTGCACGCTCAGAGTGCACCTGACCAGGCGGTGGGTCTCTCTCTCACCGG 212
|||||
QY 573 gtttgggagcgggagctgtggcccggtatcagcaaaaggtttgagcaaccaggttagaa 632
|||||
Db 211 GTCTGGGACAGAGAGCTGCTCGGGCGCGGATCCGGGGTGTGGGACTTCC-GGTAAAC 153
|||||
QY 633 actggaagtctcaggtcccagaggaatttgccttctgtctgtctgtctgtctgtctgtc 692
|||||
Db 152 ACAGACGCTCCGCTCTAGAGAAATTCGTTCCGCTGTGTCCCAAGCTCACCAGGCAG 93
|||||
QY 693 gtcaattgggcagaaaaattgttttccctcctcgtctcagcctgagcctgaattgcacctcag 752
|||||
Db 92 CTTCCTGACAGAAAAATTGGTCTTACCTGTGTGCCGAGGCTCAGGTTGCTCGTGG 33
|||||
QY 753 ggt 755
|||
Db 32 GGT 30
|||||
RNLNED 6335 bp DNA ROD 28-APR-1992
LOCUS R.norvegicus long interspersed repetitive DNA containing 7 ORF's.
DEFINITION X53581
ACCESSION X53581
VERSION X53581.1 GI:56586
KEYWORDS LINE element; long interspersed repetitive DNA.
SOURCE Norway Rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 6335)
Mohr, E.
Direct Submission
JOURNAL Submitted (21-JUN-1990) Mohr E., Institut fuer Zellbiochemie und
Klinische Neurobiologie, Universitaetskrankenhaus Eppendorf,
Universitaet Hamburg, Martinistr. 52, 2000 Hamburg 20, FRG
REFERENCE 2 (bases 1 to 6335)
Schmitz, E. and Mohr, E.
Unpublished
FEATURES
Location/Qualifiers
1..6335
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/tissue_type="liver"
/clone_lib="lambda EMBL3"
/clone="417"
670..1725
/gene="ORF1"
670..1725
/gene="ORF1"
/codon_start=1
/protein_id="CAA37644.1"
/db_xref="GI:56587"
/db_xref="SPRMBL:063303"
/translation="MVRGKRNPNSNRNODYMASSEPNSTKTNMEYPTPEKQDLVSK
SYLIMLEDFKKDVKNSLRQVAYREESQSKLKEFQENTIKOLKELKMEIAIKKEH
NETFLDENKRRQGVDTITNRIGMEERISRAEDSIEIDSTVKNVKKLLVQ
NIQIIDSRRNSRLRIIGIEESDQLGPVNFNKLIIENFPNLKREIPIGIQEAYR
TPNRLDKRNTSRHIIIVKTPNQNKERILKAVREKQVTKYKPIRITPDFSPETMKA

```

```
RRSWTDVITLREHKCPRLYPKALNSIDGETKIFHDKTKFTQYLTSTNPALQRIIN
GKAOHEASVTLLEARN"
complement(873. .1307)
/gene="ORF5"
complement(873. .1307)
/gene="ORF5"
/codon_start=1
/protein_id="CAA37648.1"
/db_xref="GI:56591"
/db_xref="SPTREMBL:Q63307"
/translation="MPMGISFFRLGKFSMILLKTTGTPLSWESSLSIPIILRFDLL
IESWSCFWTSFFRFTLSVESMISSESALEISLISILLVLMVSTAPCLFLW
FSISRVYSCSLIASISIFNSFNCLIVFWSNFRDFCDSSL"
2247. .2855
/gene="ORF2"
2247. .2855
/gene="ORF2"
/codon_start=1
/protein_id="CAA37645.1"
/db_xref="GI:56588"
/db_xref="SPTREMBL:Q63304"
/translation="NDRSKOKLNFVDRLREVMQSRLDIIYRTFPKAGYTFPSA
PRGTSKIDHIIGQTKLNRYKRIEILFCMLSDHGLKLVFNNOGRMPTIYTKLNN
LLANDLVKEEIKKEIKNFLENENEGTIPNLDMTKAVLRGLIALSACKKQERAY
VSSLFAHLKALDEKEANTPRRSRQEIILKRAEINQVETGP"
2927. .4597
/gene="ORF3"
2927. .4597
/gene="ORF3"
/codon_start=1
/protein_id="CAA37646.1"
/db_xref="GI:56589"
/db_xref="SPTREMBL:Q63305"
/translation="MRGHRECVQIKIRNEKGDITTDSEIQTILRSYKLNLSYKLE
NLQEMDFLDYQVPLKNOEIQINOLNPNIPKEIAVIGLPTKSPGPGFAEFYQ
TFIDPLILSKLPHKLETDGSLNPTYEAITLIPAPKDTTKKENFRPISLMDIDA
KILAKILANRQEHKILIIHDQVGFIPGQMGWPNIRKTNVHINLKKEQNHMIIS
LDAKADIPQHMIVKLERIGIQGYLNVKAIYKSPVANKLAGEKLEAIPKLSG
TRQCPSPFLNFVLAIRQKEVQIQIGKEVKISILFADDMIYVLDSDPKHS
TRELLKIDNF SVKARYKINSKSAFLYTKQAKEI RETTFFIIDPNIKYLGVT
LTQVKDLYNNKFTLAKETEDLRWKDLPCSWIGRINIVKAILPKAIYRNAIPI
KIYQFQKELDRICKFTWNNKPRIAKILNKRISGGTIPELQIYKAIIVIKTAW
KYPRDROIDOWNRIEDPEMHPHYGHLIFDKGAKTINKKADSIKSKWC"
4782. .5609
/gene="ORF4"
4782. .5609
/gene="ORF4"
/codon_start=1
/protein_id="CAA37647.1"
/db_xref="GI:56590"
/db_xref="SPTREMBL:Q63306"
/translation="MAYALRSRIDKWDLIKQSFCAKDTVYVTRKQPTDWNFIETNP
TDRGLLSKIYKELKLDRETNPIPKWGSSELNKEFTAEQOMAEKHLKCGSTSLVI
REMOIKTFRHLTPVPMAMIKNSGDNRCWRGCGERTLLHCWDCKLQVPFKSVWR
FLRKDLELDEPAIPLGIYPKDAPTKDCTSTFIALFTIARSWEPRCPSTEE
WIOKWYIYTMEXYSAIKNEFMKFGVKNLENIILSKLTQSQKDINGHSHLSIGY"
complement(4787. .5086)
/gene="ORF6"
complement(4787. .5086)
/gene="ORF6"
/codon_start=1
/protein_id="CAA37649.1"
/db_xref="GI:56592"
/db_xref="SPTREMBL:Q63308"
/translation="MTKDVEHFRFCSATWHSVNSLFSSEPFLIGLFLVSLRSNLF
SSLIYDIPLSVGLVGIKIFSQSGRCFVLTTSFALQKLCSFMRSHLSILDRA"
complement(5120. .5458)
/gene="ORF7"
complement(5120. .5458)
/gene="ORF7"
/codon_start=1
/protein_id="CAA37650.1"
/db_xref="GI:56593"
/gene="ORF7"
complement(5120. .5458)
/gene="ORF7"
/codon_start=1
/protein_id="CAA37650.1"
/db_xref="GI:56593"
```

```
/db_xref="SPTREMBL:Q63309"
/translation="MYHIFCIHSSVEGHLSFOLLALIIKAMNIYVHVSFLVVGASF
GYMPKRGISAGSSGSSMSNFRNLQTFQNGCTSLQSHQWRSVPLSPHPRQLLSPEF
:LIILALIGVR"
BASE COUNT 2522 a 1381 c 1245 g 1187 t
ORIGIN
```

```
Query Match 15.88; Score 172; DB 88; Length 6335;
Best Local Similarity 64.18; Pred. No. 4.7e-43;
Matches 418; Conservative 0; Mismatches 170; Indels 64; Gaps 8;

Qy 1 ttgtctctgg-gttac-ttgtttccattcttgacagtggtttgac-ctctatacgcc 58
Db 665 TTATCTCTAGTGTACTTTGTTCTGCTATTCTGACAGTGTGTAGACTGCTCTATAAGCC 606

Qy 59 tgtgttcagagtgctgttagaacctatttctctgtttctttcttcagccagttacaggaaca 118
Db 605 TGTGTGTCAGAGTGTGTAGACTGTTTTC---TCTCTTCAGTCCGTTATGGGACA 549

Qy 119 gagtgttctactgtcagatgttagctgttctctgtccactgactttcaa----- 167
Db 548 GAGTGTCTCTCTTCGGGCGTGTAGTTTTCCTCTCTGAGGTCCTCAGCTGTCTCTGTG 489

Qy 168 -----gctgtctctgtgtgcagga 186
Db 488 GGCTGTGTCTTGTAGTTCACAGCGGGGTGCTGCCAAGGGGCTCTCTGTGGGGCAGCA 429

Qy 187 accagaagggcctgtccctacttctactgg-gccctacgcacaggggcccctagatggtg 245
Db 428 ACCAGGAGAGACTGTCCGCCGCCCTCCAGAGAGCTTCAGTGACACAGGGTTCAGATGGCC 369

Qy 246 ctagggttttccctctagagcctgaatgtggg--cagagagtagtctctctctggtttcc 303
Db 368 TTTGGTGTCTTCTCTGGGCTCGAGATGTATGTACAGAGAGACAGTCTCTTCTGTTTCC 309

Qy 304 taggtatgtctccctctgaagtgtagctctctccctccatgggatacgggtgcagga 363
Db 308 AAGCTTGCTGCTCTCTGAAGGTTTCAGCTCTCCCTCCCATGGGATTTGGGTGGCAGAGA 249

Qy 364 gctgtttgaccag--gtcctctcaaatccgggtgcagtcctgcacgcaggtccctgtacg 421
Db 248 ACTGTTTATCCGGTCTGTTTCTCCTCAGGTTCGGCGGGTCTCTCAGGACAGGGGTCTGACG 189

Qy 422 ttgctgtgtgaatctccgcacccagagc--accagaagttctctcttgggccaagga 479
Db 188 CTCCTGGGCCCTTCCCGCCAGCGAACCAGAGGCGCTTATACAGTTTCTCTTGGGCCAGGA 129

Qy 480 tgtgggcaaggtgggcagaaagtggaatctctctccctagcgtctcagagattgccct 539
Db 128 TGTGGCAGGGGTGGCAGTGTGTGGTGTCTCTCCGCTCTGACGCTCAGAGGTCAGAGTGCCCA 69

Qy 540 cactctctggggaatccgctctctctctccacaggggtttgggagcagggagctg 591
Db 68 CCGTACCAGGCGGTGAGGTCTCTTTCTCACGGGTCTCTGGGAGCAGAGAGCTG 17

RESULT 12
AF000202/c
LOCUS
DEFINITION
Rattus sp. T-612 retrotransposon mlv12-rm6 5'UTR and putative RNA
binding protein gene, partial cds.
ACCESSION
AF000202
VERSION
AF000202.1
KEYWORDS
GI:2599312
SOURCE
Rattus sp. T-612.
ORGANISM
Rattus sp. T-612
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1117)
Cabot, E.L., Angeletti, B., Usdin, K. and Furano, A.V.
Rapid evolution of a young LI (LINE-1) clade in recently speciated
```

Mon Nov 6 10:14:43 2000

## RESULT 13

AF000203/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

repeat\_region

5'UTR

CDS

Rattus taxa

J. Mol. Evol. 45 (4), 412-423 (1997)

79472947

2 (bases 1 to 1117)

Cabot, E.L. and Furano, A.V.

Direct Submission

Submitted (18-APR-1997) LNCB-NIDDK, NIH, Bldg. 8, Room 203, 8

Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA

Location/Qualifiers

1. .1117

/organism="Rattus sp. T-612"

/isolate="T-612"

/db\_xref="taxon:68896"

/note="from the collection of Institut des Sciences de

1'Evolution, Montpellier II; collected from the island of

Ceram, Indonesia and identified by Djoko Iskandar as

Rattus moluccarius; cranial morphology suggests similarity

to Rattus rattus satarae"

1. .&gt;1117

/note="m1v12-rm6"

/rpt\_family="LINE 1 non-LTR retrotransposable element"

/rpt\_type="dispersed

1. .733

734. .&gt;1117

/note="ORF 1"

/codon\_start=1

/product="putative RNA binding protein"

/protein\_id="AAC53497.1"

/db\_xref="GI:2599313"

/translation="MARGKRRNPSPNODCMPSPPSPPTKINMEYPTNPKQDLVSK

SYLIMLEDFKDMTLRETQIINKQVAYREWQKSLKFEQNTIKQLKELMEIE

AIKKEHMETTLDIENOKRQGVDTs"

389 a 295 c 277 g 156 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

0; Mismatches

137; Indels

9; Gaps

4;

QY 168 gctgtctgtgtgcaggaacacgaaggccgtccctacttactgtgcccctacgca 227

DB 451 GCTCTCTCCAGCGGACAGACACGAGAGACCTGTGGCGCCCTTCGGGAGCTTCAGTGC 392

QY 228 cagggggcctagatggtgtcgtggttttctctctagagctgaatgtg-gcagagag 285

DB 391 ACCAGGTTCCAGATGGTCTTTGGCTTTTCTCTGGCGTCCGAGATGTGTGCAGGGAG 332

QY 286 tagtctctctgttctctaggtgtgtcttccctctgaaggtctagctctccctccat 345

DB 331 CAGTCTCTCTGGTTCCAGGCTGTGTCTGCTCTCTGAAGGTTTCAGCTCTCCCTCCAC 272

QY 346 gggataggggtgcagggagctgtttt---gaccaggtctctcctcctcctcctcctcctcct 401

DB 271 GGGATTGGTGCAGAGAACTTTTATCCGGTCTGTCTTTCAGGTTCTGGCGGTCTCT 212

QY 402 ggaacagag-gtctctgtgctgtcgtgtcgtcctcctcctcctcctcctcctcctcctcct 458

DB 211 CAGGCGCAGAGGCTCTCTGGCGCTCTCTGGCGCTCTCTGGCGCTCTCTGGCGCTCTCT 152

QY 459 agttctctcttggtcgaagatgtggcgaaggtgggagagtggaatctcctcctcctcctcct 518

DB 151 AGTTTCTCTTGGGCCAGGAGGTGGGCGAGGGGTGAGCAGATGTGTGGTGTCTCTTCCGCT 92

QY 519 ctgagctctcagagattgcctcactctctgtggcaatccctctctcctcctcctcctcctcct 578

DB 91 CTGACGCTCAGGAGTGCCCACTGACACAGGCGGTGGGTCTCTCTCTACCGGGTCTGG 32

QY 579 gacagagagctg 591

DB 31 GAGCAGAGAGCTG 19

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

Rattus sp. T-612

```

Db 221 TCAGCGCAGAGTCTCGCGCTCTCTCCCTCCCAACAGCGGAGCCAGAGGCTTATA 162
QY 458 aagtttctcttggccaaagatgtggccaaagtgggcagaagtggcaatctctctgc 517
Db 161 CAGTTTCTCTTGGCCAGGAGTGTGGCAGGGGTGAGCAGTGTGGTGTCTCTCCGC 102
QY 518 ctagctctcagagattgacctctctctggcgaatccgctctctcttccacagggttg 577
Db 101 TCTGACGCTCAGGAGTCCCATCTGACCGCGGTGGGCTCTCTCTACCGGGCTG 42
QY 578 ggagcaggagcttggccgggtatcaggcgaaggtttgag 618
Db 41 GGAGCAGAGAGCTGCTGGCGGCGGATCCGCGGTGTGG 1

RESULT 14
AF000200/c
LOCUS
DEFINITION
Rattus sp. T-612 DNA ROD 11-DEC-1997
Binding protein gene, partial cds.
ACCESSION
AF000200
VERSION
AF000200.1 GI:2599308
KEYWORDS
Rattus sp. T-612.
SOURCE
Rattus sp. T-612.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1127)
Cabot, E.L., Angeletti, B., Usdin, K. and Furano, A.V.
Rapid evolution of a young L1 (LINE-1) clade in recently speciated
Rattus taxa
J. Mol. Evol. 45 (4), 412-423 (1997)
MEDLINE
97472947
REFERENCE
2 (bases 1 to 1127)
Cabot, E.L. and Furano, A.V.
Direct Submission
JOURNAL
Submitted (18-APR-1997) LMCB-NIDDK, NIH, Bldg. 8, Room 203, 8
Center Dr. MSC 0830, Bethesda, MD 20892-0830, USA
FEATURES
Location/Qualifiers
1..1127
/organism="Rattus sp. T-612"
/isolate="T-612"
/db_xref="taxon:68896"
/note="from the collection of Institut des Sciences de
Ceran, Indonesia and identified by Djoko Iskandar as
Rattus moluccarius; cranial morphology suggests similarity
to Rattus rattus satarae"
1..>1127
/note="mlv12-rm3"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type="dispersed"
1..743
744..>1127
/codon_start=1
/note="ORF 1"
/product="putative RNA binding protein"
/protein_id="AAC53495.1"
/db_xref="GI:2599309"
/translation="MARGKRNPNSNRNQCMPSPSPKTNMYPNTPEKQDLVSK
SYLLIMLEDFKKDMTLEQEIINKQVAYREWKSLKEFOENTIKQMKELKWEIE
AIKHEMTTLDIENKKROGAVDTIS"
BASE COUNT 390 a 308 c 273 g 156 t
ORIGIN

Query Match 13.78; Score 149.6; DB 88; Length 1127;
Best Local Similarity 66.8%; Pred. No. 4.8e-36;
Matches 290; Conservative 0; Mismatches 134; Indels 10; Gaps 5;
Qy 168 gctgtctgtgtgcaggaaacagaggcctgcccctacttacttgcg-gccctacgc 226

```

```

Db 461 GCCTCTGTCAGCGGAGCAGCAACAGAGACCTGTGCGCCCTTCCGGAGCTTCAGTGC 402
QY 227 acagggggcctagatggtgctaggtgtttctctctagagcctgaaatgtgg--gcagaga 284
Db 401 ACAGGGTTCAGATGGTCTTTGGCTTTTCTCTGCGTCCGAGATGTGTGTGAGGGA 342
QY 285 gtagtctctctgttctctaggtatgtctctctctctctctctctctctctctctctca 344
Db 341 CGAGTCTCTCTGGTTCCTCCAGCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 282
QY 345 tgggatatgggtgcaggagagctgttt----gaccaggtctctctcaaatccgggtgcagtc 400
Db 281 CGGGATTGGGTGCAGAGAACTGTTTATCCGGTCTGTTTCTTCAGGATCCGGTGGTGC 222
QY 401 tggaccgcag-gctctgtagctgtcgtcgaatctctctctctctctctctctctctctcc 457
Db 221 TCAGCGCAGAGTCTCTCCGCTCTCTGGGCTCTCTCCCAACAGCGGAGCCAGAGGCTTATA 162
QY 458 aagttctcttggccaaagatgtggcgaagtgggcagaagtggcaatctctctctgc 517
Db 161 CAGTTTCTCTTGGCCAGGAGTGTGGCAGGGGTGAGCAGTGTGGTGTCTCTCTCCGC 102
QY 518 ctagctctcagagattgacctctctctggcgaatccgctctctcttccacagggttg 577
Db 101 TCGGACGCTCAGGAGTCCCATCTGACCGCGGTGGGCTCTCTCTCTCTCTCTCTCT 42
QY 578 ggagcaggagagctg 591
Db 41 GGAGCAGAGAGCTG 28

RESULT 15
RN38NLVII/c
LOCUS
DEFINITION
Rattus norvegicus L1 retrotransposon mlv12-rn38, 5'UTR and putative
RNA binding protein 1 gene, partial cds.
ACCESSION
U87604
VERSION
U87604.1 GI:2047334
KEYWORDS
1 of 2
SEGMENT
Norway rat.
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1113)
Hayward, B.E., Zavanelli, M. and Furano, A.V.
Recombination creates novel L1 (LINE 1) elements in Rattus
norvegicus
Genetics (1997) In press
REFERENCE
2 (bases 1 to 1113)
Hayward, B.E., Good, B., Cabot, E. and Furano, A.V.
Direct Submission
JOURNAL
Submitted (29-JAN-1997) LMCB/NIDDK, NIH, 8/203, 8 Center Dr. MSC
0830, Bethesda, MD 20892-0830, USA
FEATURES
Location/Qualifiers
1..1113
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/note="L1mlvrn38b"
<1..729
/note="mlv12-rn38"
/rpt_family="LINE 1 non-LTR retrotransposable element"
/rpt_type="dispersed"
730..>1113
/note="ORF 1"
/codon_start=1
/product="putative RNA binding protein 1"
/protein_id="AAB53049.1"
/db_xref="GI:2047337"
/translation="MARGKRNPNSNRNQCMPSPSPKTNMYPNTPEKQDLVSK"
5'UTR
repeat_region
CDS

```



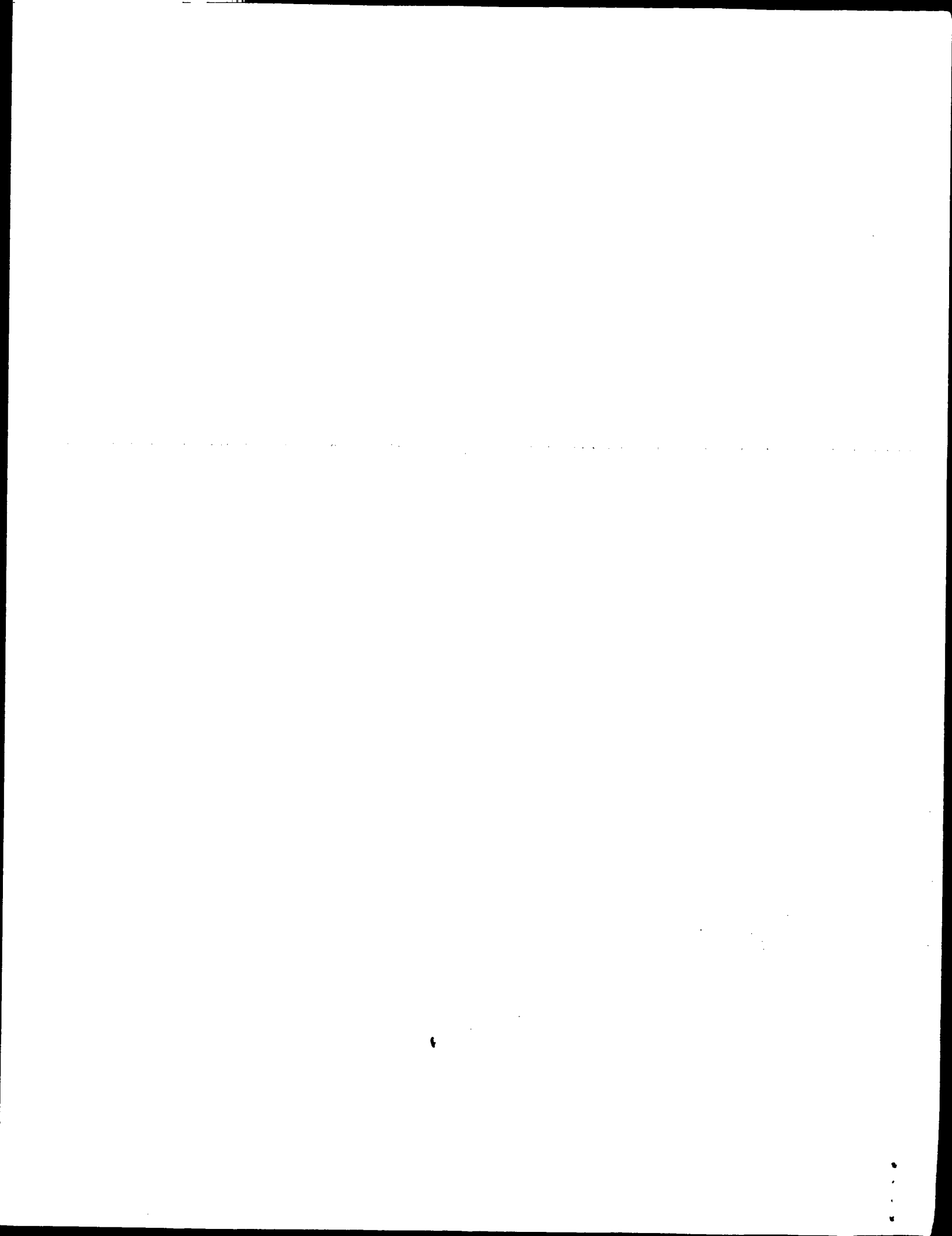
SYLIMLEDFKDMNTLREAGENINKQVAYRESQSLKEFOENTIKQLKELMEIE  
AIKKEHMETTLIDENQKROGAVDTS

BASE COUNT 390 a 300 c 272 g 151 t  
ORIGIN

Query Match 13.5%; Score 147.6; DB 88; Length 1113;  
Best Local Similarity 63.7%; Pred. No. 2.1e-35;  
Matches 275; Conservative 0; Mismatches 149; Indels 8; Gaps 3;

Qy	168	gctgtctctgtgaggaaccagagggccctgctccctacttctacttg-gccccctacgc	226
Db	447	GCTCTCTGACGCGCACACACAGGAGACCTGTGCCGCCCTTCGGGAGCTTCAGTGC	388
Qy	227	acagggggcctagatggtgctaggtgttttccctctagagcctgaaatgtgg--gcagaga	284
Db	387	ACCAGGTTCCAGATGCTCTTTGGCTTTTCCCTCTGGCTGCCGATGTGTGCAGGGA	328
Qy	285	gtagtctctctgtttccctaggtatgtctccctctctgaagctctagctctccctcca	344
Db	327	GCAGTCTCTTCCGGTTTCCAGGCGCTGTGCTCTCTGAAGCTTTAGCTCTCCCTCCCA	268
Qy	345	tggatatgggtcagggagctgttgaccag-----gtcctctcaaatccgggtgcagt	399
Db	267	CGGATTTGGGTGCAGAGACCTGTTTATCCGGTCTGTCTTCCTTCAGGTTCCGGTGGTGC	208
Qy	400	ctggaccgcaggtcctctgtgctgtgcctgctgcaatcttcccgcacccagagggacccaa	459
Db	207	TCTGCGAGGGTCTCTGCGCTCTCTGGGCCCTCCGCCACGGGAGCCAGAGCCCTTATACA	148
Qy	460	gtttctcttggccaaaggtatgtgggcaaggtgggcagagtggaatctctctctgccc	519
Db	147	GTTCCTCTTGGGCCAGGGATGTGGGAGGGGTGAGCAGTGTGGGAGGTCTCTCCGCTC	88
Qy	520	tagcgtctcaggttgcctcactctctgggcaatcccgctctctctctccacaggggtttggg	579
Db	87	TGCACCTTCAGGAGTGCCCACTGCACAGGGGGTGGGTCTCTCTCTCACCGGGTCTGGG	28
Qy	580	agcagggagctg 591	
Db	27	AGCAGAGAGCTG 16	

Search completed: November 4, 2000, 01:24:47  
Job time: 26728 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 00:31:40 ; Search time 254.92 seconds  
(without alignments)  
1606.278 Million cell updates/sec

Title: US-09-101-423A-6  
Perfect score: 1090  
Sequence: 1 tggctctgtgttacttgt.....cctagcacagtcacccaca 1090

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues  
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36:\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1090	100.0	1090	18 T72790	Metastasis inducin
2	224.6	20.6	44576	21 Z61522	Cosmid CVO14 conta
3	121.8	11.2	2225	21 Z94380	Rat haem oxygenase
4	97.6	9.0	437	15 Q65636	P35 sequence of hu
5	96	8.8	437	15 Q65637	Rat retrotransposo
6	34.2	3.1	282	20 V69576	Human ZCHEMO-8 deg
7	34.2	3.1	307	21 A10207	Rat liver toxicolo
8	33.4	3.1	1818	19 V17342	Coding sequence fo
9	32.8	3.0	570	20 Z08665	Human interleukin
10	32.8	3.0	570	20 X17786	Human interleukin
11	32.4	3.0	10732	21 A10594	Gene encoding a su
12	31.6	2.9	3835	20 X87772	Human mucin gene M

c 13	31.4	2.9	1295	20 X20325	Borrelia burgdorfe
14	31.4	2.9	8438	15 Q73500	DNA encoding pseud
c 15	31.2	2.9	1026	21 Z37262	ScRF coding sequen
c 16	31.2	2.9	1067	21 A16686	Human secreted pro
c 17	31.2	2.9	2728	18 T58853	cDNA encoding high
c 18	31.2	2.9	2728	19 V28861	Rat proline transp
c 19	31	2.8	2114	14 Q31236	Plant NADPH cytoch
20	31	2.8	15328	16 Q81139	HPLA2-8 gene. Hom
c 21	30.8	2.8	321	14 Q36580	Light chain of rat
c 22	30.8	2.8	698	21 A01621	Human colon cancer
c 23	30.6	2.8	411	19 V10405	Human myelin P2 pr
c 24	30.6	2.8	522	19 V62400	Human sonic hedgeh
c 25	30.6	2.8	1425	16 Q91639	Human sonic hedgeh
c 26	30.6	2.8	1425	20 X25622	Human sonic hedgeh
c 27	30.6	2.8	1425	20 X25103	Human sonic hedgeh
c 28	30.6	2.8	1425	20 X07276	Human sonic hedgeh
c 29	30.6	2.8	1425	20 X16187	Human Shh hedgehog
c 30	30.6	2.8	1425	21 Z52262	Human sonic hedgeh
c 31	30.6	2.8	1576	19 V18403	Human mutated soni
c 32	30.6	2.8	1576	19 V18404	Human mutated soni
33	30.6	2.8	2150	16 Q79581	Lymphocyte chemoat
34	30.6	2.8	2150	18 T49000	Lymphocyte chemoat
35	30.6	2.8	2151	17 T33895	Escherichia coli f
c 36	30.6	2.8	2599	21 Z56382	Lymphocyte chemoat
37	30.6	2.8	3175	21 A35226	Human adenosine re
38	30.6	2.8	3568	21 A35227	Human adenosine re
39	30.6	2.8	9096	18 T88015	Human interleukin
40	30.4	2.8	3589	18 V74576	Staphylococcus aur
41	30.2	2.8	1892	21 Z64956	Membrane-bound pro
42	30	2.8	450	19 V45445	Human chemokine ZS
43	30	2.8	3800	19 V36329	Human LBRGAL gene
44	30	2.8	3805	20 Z09524	Human Apo E genomi
45	30	2.8	3805	20 Z09526	Human Apo E genomi

## ALIGNMENTS

RESULT 1  
T72790  
ID T72790 standard; DNA; 1090 BP.  
XX  
AC T72790;  
XX  
DT 22-SEP-1997 (first entry)  
XX  
DE Metastasis inducing DNA C20.  
XX  
KW Metastasis-inducing DNA; Met-DNA; cancer; tumour; diagnosis;  
KW osteopontin; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9725443-A1.  
XX  
PD 17-JUL-1997.  
XX  
PF 10-JAN-1997; 97WO-GB000074.  
XX  
PR 10-JAN-1996; 96GB-0000470.  
XX  
PY (UYLI-) UNIV LIVERPOOL.  
XX  
PA Barraclough BR, Rudland PS;  
XX  
PI WPI; 1997-372878/34.  
XX  
DR New isolated metastasis-inducing DNA - used to develop products to  
PT identify and treat patients at risk from metastatic tumours  
XX  
PS Claim 13; Page 27; 38pp; English.  
XX  
CC Metastasis-inducing DNA (Met-DNA) fragments C2, C5, C6, C9, C12

CC and C20 (72785-90) are entirely novel short stretches of human  
CC regulatory DNA capable of inducing metastasis. They were  
CC identified using a method for detecting Met-DNA that involves  
CC transferring human DNA from a malignant, metastatic cancer cell  
CC (in this case, breast cancer) into a cell line (pref. rat Rama 37)  
CC that produces only benign, non-metastasing tumours when injected  
CC into a syngeneic animal, selecting those animals having metastasing  
CC syngeneic animal, and recovering the Met-DNA from them. The isolated  
CC Met-DNAs can be used to develop products to identify and treat  
CC patients at risk from metastatic tumours.

XX Sequence 1090 BP; 224 A; 277 C; 275 G; 314 T; 0 other;

Query Match 100.0%; Score 1090; DB 18; Length 1090;  
Best Local Similarity 100.0%; Pred. NO. 0;  
Matches 1090; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtctctgtgttactgttttccattttctgacagtggtttgacatttatacagctg 60  
DB 1 ttgtctctgtgttactgttttccattttctgacagtggtttgacatttatacagctg 60

QY 61 ttgtctcagagtgctgtagacatttttctgttttttttttttttttttttttttttttt 120  
DB 61 ttgtctcagagtgctgtagacatttttctgttttttttttttttttttttttttttttt 120

QY 121 gtgtttactgtcagatgtgttagctgttcttccactgacttttcaagtgctctctgtgt 180  
DB 121 gtgtttactgtcagatgtgttagctgttcttccactgacttttcaagtgctctctgtgt 180

QY 181 gcaggaaacagagggcctctcactatttcttactgttggccctacgcacagggggcctaga 240  
DB 181 gcaggaaacagagggcctctcactatttcttactgttggccctacgcacagggggcctaga 240

QY 241 ttgtctcagagtgctgttctcactagacgtgaaatgtggcagagagtagtctctctgtgt 300  
DB 241 ttgtctcagagtgctgttctcactagacgtgaaatgtggcagagagtagtctctctgtgt 300

QY 301 tcttagtgatgttttccctctgaaagtgtagctcttcccttccatgggatatgggtgag 360  
DB 301 tcttagtgatgttttccctctgaaagtgtagctcttcccttccatgggatatgggtgag 360

QY 361 ggagctgtttgacagagtgctctcactgacgtgagtgagtgacgagagtgctctctgtag 420  
DB 361 ggagctgtttgacagagtgctctcactgacgtgagtgagtgacgagagtgctctctgtag 420

QY 421 ctgctctgtcgaattcttccctcagcagagggcagcccaagtttctcttgggccaagat 480  
DB 421 ctgctctgtcgaattcttccctcagcagagggcagcccaagtttctcttgggccaagat 480

QY 481 gtgggcaaaaggtggcagagtgagtggaattcttctcctgacctagctcagagtgacctc 540  
DB 481 gtgggcaaaaggtggcagagtgagtggaattcttctcctgacctagctcagagtgacctc 540

QY 541 acttctgggcaatccgctctcttccacagaggttttggagcagagtgagtgagtgagtgag 600  
DB 541 acttctgggcaatccgctctcttccacagaggttttggagcagagtgagtgagtgagtgag 600

QY 601 atcaggaacaggtttgaggaacacaggttagaactggaagtgtcaggttccagaggaatt 660  
DB 601 atcaggaacaggtttgaggaacacaggttagaactggaagtgtcaggttccagaggaatt 660

QY 661 ttgctctgtgtgtcctgagtcacacagcagaggttcaacttgagcagagaaaaattgttttc 720  
DB 661 ttgctctgtgtgtcctgagtcacacagcagaggttcaacttgagcagagaaaaattgttttc 720

QY 721 cctcgtgtcagagcctgaattggaactcaggttgggttttcaagtgtagcttgagtgagaa 780  
DB 721 cctcgtgtcagagcctgaattggaactcaggttgggttttcaagtgtagcttgagtgagaa 780

QY 781 gtatggttt 840  
DB 781 gtatggttt 840

DB 781 gtatggttt 840  
QY 841 ttgttcttt 900  
DB 841 ttgttcttt 900  
QY 901 atcacatttctgtcactgacgtcagctcatttcttttttttttttttttttttttttttttt 960  
DB 901 atcacatttctgtcactgacgtcagctcatttcttttttttttttttttttttttttttttt 960  
QY 961 ctgcgcagcagaaacacacgctagggacatacgaatcttctgagccaaactttat 1020  
DB 961 ctgcgcagcagaaacacacgctagggacatacgaatcttctgagccaaactttat 1020  
QY 1021 tgaattcttaaggagagcccgccgacccgactggcggttttatatacacctagcacag 1080  
DB 1021 tgaattcttaaggagagcccgccgacccgactggcggttttatatacacctagcacag 1080  
QY 1081 tgcattccaca 1090  
DB 1081 tgcattccaca 1090

RESULT 2  
261522/c  
ID Z61522 standard; DNA; 44576 BP.

XX Z61522;  
XX 19-JUN-2000 (first entry)

XX Cosmid CVO14 containing rat vasopressin and oxytocin locus.

XX Oxytocin expressed sequence tag; 5'-OT EST; obesity; fertility; male;  
XX transgenic animal; human late onset obesity; late onset visceral obesity;  
XX male infertility; wasting; anorexia; cachexia; malabsorptive state;  
XX catabolic state; inflammatory condition; Crohn's disease; AIDS wasting;  
XX burn; cancer; bone disease; vasopressin; oxytocin; ss.

XX Rattus sp.

XX WO200009686-A1.

XX 24-FEB-2000.

XX 12-AUG-1999; 99WO-GB02658.

XX 12-AUG-1998; 98GB-0017566.

XX 06-MAY-1999; 99GB-0010522.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Robinson ICAF, Stoye JP, Flavell D, Wells SE, Le Tissier P;

XX WPI; 2000-224331/19.

XX New anti-obesity polypeptide useful for treating obesity or infertility  
XX in mammals

XX Claim 9; Page 129-154; 162pp; English.

XX The present sequence represents cosmid CVO14, which contains the rat  
XX vasopressin and oxytocin locus. The specification describes 5'-OT-EST  
XX (oxytocin expressed sequence tag) polypeptides. The 5'-OT EST gene is  
XX involved in the control of obesity and fertility in males. 5'-OT EST  
XX nucleic acids are useful for producing transgenic animals. The  
XX transgenic animals created serve as a model for human late onset  
XX obesity and other related disorders and are also used for identifying  
XX the genetic cause of obesity. Compounds which modulate 5'-OT EST  
XX expression or activity are useful in the treatment of modulating of  
XX late onset visceral obesity or male infertility particularly in the  
XX disorders related to these conditions such as wasting, or anorexia,  
XX or cachexia associated with prolonged illness, or malabsorptive states

us-09-101-423a-6.rng

Mon Nov 6 10:14:44 2000

CC or catabolic states associated with other diseases such as inflammatory  
 CC conditions, Crohn's disease or AIDS wasting, or burns, or cancer, or  
 CC bone disease.  
 XX Sequence 44576 BP; 12157 A; 10993 C; 10857 G; 10569 T; 0 other;  
 SQ

Query Match 20.6%; Score 224.6; DB 21; Length 44576;  
 Best Local Similarity 64.0%; Pred. No. 2.9e-62;  
 Matches 515; Conservative 0; Mismatches 224; Indels 56; Gaps 9;

Qy 1 ttgtctctgtgttac-ttatttccattctgacagtggtttgac-cttctatacgc 58  
 Db 14487 TTATCTCTAGTGTACTTGTCTGCTATTCTGACAGTGTAGACTGCTCTATAAGCC 14428  
 Qy 59 tgtgttcagagtgctgtgtagacatttctctgtttttctttcagccagttacagaa 118  
 Db 14427 TGTGTGTGAGGAGTGTCTAGACCTGTTTCC---TCTCTTTCAGTCCGTTATGGGACA 14371  
 Qy 119 gsgtcttactgtcagatgttagctgttctctcactgactttcaa----- 167  
 Db 14370 GAGTGTCTGCTTTCGGGCGTGTAGTTTCTCTCTGACAGTCTTCAGCTGTTCTGTG 14311  
 Qy 168 -----gctgtctctgtgtgca 186  
 Db 14310 GGCTGTGCTGTGAGTTCACAGCGGGGTGCTGCCAAGGGCTCTCTGTGGGGCAGCA 14251  
 Qy 187 accagaagggcctgtctctacttctactgg-gccctacgcacagggggcctagatggtg 245  
 Db 14250 ACCAGAAGACCTGTGCGGCCCTTCCAGGAGCTTCAGTGCACCGAGGTTCAGATGCC 14191  
 Qy 246 ctagggtttctctctagagcctgaatg--tgggcagagagtagtctctctctgtttcc 303  
 Db 14190 TTGTGTTTCTCTGCGGTCGAGATGTATGTACAGAGACGACTCTCTCTGTGTTC 14131  
 Qy 304 taggtatgtctctctctgaaggttagctctctctctctctctctctctctctctct 363  
 Db 14130 AAGGCTGTGTGCTCTCTGAAGGTTTCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14071  
 Qy 364 gctctttgaccag-gtctctctcaaatccgggtgtagtctgacgcagggctctctgtagc 421  
 Db 14070 ACTGTTTATCCGCTCTGTTTCCCTCAGGTTCCGGCGGTGTCTCAGGCGAGGGTCTGAGC 14011  
 Qy 422 ttgctgtctgaatctctccgcaccagagc--accaagttctcttggcccaagga 479  
 Db 14010 CTCTCTGGCCCTCCCCACAGGAACCCAGAGGCTTATACAGTTCTCTCTTGGCCAGGGA 13951  
 Qy 480 tgtgggcaaggttgagcagaagtggaatctctctgcccctagcgttctcagagattgcct 539  
 Db 13950 TGTGGGCAAGGGTGGGAGTGTGTGGTGTCTCTCTCCGCTCTGAGGCTCAGAGTGCCCA 13891  
 Qy 540 cactctgtggcaatccgctctctctctctccacaggggtttggagcagagagcttgcccg 599  
 Db 13890 CCTGACACAGCGGTGAGGTCTCTCTCTCAGGGTCTTGGAGCAGAGAGTGTCTG--CG 13833  
 Qy 600 taccagcaaggttgtaggcaacacttagaactgaaagtgtcaggtctccagaggaat 659  
 Db 13832 GCCGGGATCCCGGGTGTGGGACTTCGGCAACACAGAGAGTGCCTAGATGAAC 13773  
 Qy 660 ttgtctgtgtgtctgtctgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 719  
 Db 13772 TGTGCTCTGTGTGTCCTCAATCTCAGGAGGAGTCTCTCTTTCAGCAGACAAGTGTGCT 13713  
 Qy 720 cccctcgttctcagcctgaagttg 744  
 Db 13712 ACCTGTGTCCTCGAGGCTCAAGTTG 13688

RESULT 3  
 ID Z94380  
 ID Z94380 standard; cDNA; 2225 BP.  
 XX  
 AC Z94380;

XX 03-JUL-2000 (first entry)  
 DT Rat haem oxygenase HO-3 cDNA.  
 DE  
 XX  
 XX HO-3; haem oxygenase; rat; heart transplant; preservation;  
 KW stenosis; resensitis; myocardial infarction; myocarditis;  
 KW vascular injury; gene therapy; antisense therapy; ss.  
 XX  
 XX Rattus rattus.  
 XX Location/Qualifiers  
 Key 1062..1934  
 CDS /\*tag- a  
 WO200012118-A2.  
 XX  
 XX 09-MAR-2000.  
 XX 27-AUG-1999; 99WO-US19823.  
 XX 28-AUG-1998; 98US-0098377.  
 XX 25-FEB-1999; 99US-0121946.  
 XX (HARD ) HARVARE COLLEGE.  
 XX (LEEM/) LEE M.  
 XX (PERR/) PERRELLA M A.  
 XX (YETS/) YET S.  
 XX Lee M, Perrella MA, Yet S;  
 XX WPI; 2000-256492/22.  
 XX P-PSDB; Y79283.  
 XX Inhibiting cardiomyocyte death by administering a haem oxygenase is  
 PT useful to treat cardiovascular injury and to preserve a heart for  
 PT transplant purposes  
 XX  
 PS Disclosure: Page 19-21; 27pp; English.  
 XX  
 XX The present sequence is that of cDNA encoding rat haem oxygenase  
 CC HO-3 (see Y79283). The invention features methods of minimizing  
 CC myocardial damage by salvaging hypoxic myocardial tissue before it  
 CC becomes irreversibly injured. A method of inhibiting  
 CC cardiomyocyte death in a mammal, e.g. a human, who has suffered a  
 CC myocardial infarction or who has myocarditis, is carried out by  
 CC administering to the myocardium a HO polypeptide, preferably  
 CC selected from HO-1, HO-2 or HO-3, or by administering DNA encoding  
 CC an HO. Cardiomyocyte death is inhibited in vitro by contacting  
 CC cardiomyocytes with HO or DNA encoding HO. For example, isolated  
 CC myocardial tissue, such as a donor heart to be used in  
 CC transplantation, is preserved for up to 24 hr by bathing or  
 CC perfusing with a solution containing HO or DNA containing HO.  
 CC Vascular smooth muscle cell proliferation is inhibited by  
 CC administering HO to injured vascular tissue within 24 hr, or up to  
 CC 1 wk, after injury.  
 XX  
 XX Sequence 2225 BP; 539 A; 509 C; 552 G; 625 T; 0 other;  
 SQ

Query Match 11.2%; Score 121.8; DB 21; Length 2225;  
 Best Local Similarity 62.9%; Pred. No. 2e-29;  
 Matches 222; Conservative 0; Mismatches 127; Indels 4; Gaps 2;

Qy 168 gctgtctctgtgtcaggaaccagagggcctctccctacttacttctgagccctacgca 227  
 Db 516 gctctctgtcaggggcaacacacaggaagacctgtgcggccctccggagcttcagtgca 575  
 Qy 228 caggggctcagagtggtgtctaggtgttttctctctagagcctgaattgtgg--cagagag 285  
 Db 576 ccaggggttcagagtggt 635  
 Qy 286 tagtctctctgt 345

```

Query Match      8.8%; Score 96; DB 15; Length 437;
Best Local Similarity 81.0%; Pred. No. 2.1e-21;
Matches 136; Conservative 0; Mismatches 30; Indels 2; Gaps 2;

YY 1 ttgtctctggcggttac-ttgtttttcccatcttgacagtgttgac-cttctatagcc 58
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
bb 312 TTATCTCTAGTGTACTTTTGTCTGCTATTTCTGACAGTGGCTAGACTGCCCTATAAGCC 253
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

YY 59 tgtgtctcagcagtgctgtagacctatttccctgtttcttcagccaggttacagggaaca 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
bb 252 TGTGTCTCAGGAGTCTGTAGACCTGTGTTTCTGTTTCTTCATTGACGCTGTATGCGGACAA 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

YY 119 gagtgtctctactgtccagatgtgagctgttctctgtctctctccactgaacttca 166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
bb 192 GAGTGTCTCTGCTTCGGCGGTGAGTTTTTCTCTCTCTACAGGTTCTCA 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
D 69576/c
D V69576 standard: cDNA; 282 BP.
X X V69576;
C C
X X
X X
T T
T T
T T
E E
X X
Human ZCHEMO-8 degenerate cDNA.

Chemokine; ZCHEMO-8; human; pathological condition; infection; cancer;
autoimmune disorder; immunodeficiency; myeloproliferative; wound healing;

```

us-09-101-423a-6.rng

MORF NOV 6 10:14:44 2000

KW transplant; progenitor cell; HIV infection; AIDS; chemotherapy;  
 KW radiation therapy; T cell; macrophage activation inhibitor; B lymphocyte;  
 KW chronic inflammatory disease; infective disease; diagnosis; detection;  
 KW drug screening; gene therapy; ds.  
 XX Homo sapiens.  
 XX WO9854326-A1.  
 XX 03-DEC-1998.  
 XX 19-MAY-1998; 98WO-US10329.  
 XX 29-MAY-1997; 97US-0047860.  
 XX (ZYMO) ZYMOGENETICS INC.  
 XX Sheppard PO;  
 XX WPI; 1999-059841/05.  
 XX New isolated human beta-chemokine, ZCHEMO-8 - used to develop  
 XX products for treating e.g. ischaemia, reperfusion, wound healing,  
 XX autoimmune diseases, inflammation, asthma or infections  
 XX Claim 23; Page 104; 131pp; English.

XX This sequence encodes a human beta chemokine, ZCHEMO-8. Altered levels  
 XX of ZCHEMO-8 may be indicative of pathological conditions, including  
 XX infections, cancer, myeloproliferative disorders, autoimmune disorders and  
 XX immunodeficiencies. The ZCHEMO-8 polypeptides can be used, e.g. to  
 XX reduce the damage in ischaemic and reperfusion injuries, in a wound  
 XX healing regime to stimulate an infiltration of immune cells (e.g.  
 XX monocytes, neutrophils, T lymphocytes or basophils) to a wound site to  
 XX facilitate healing. ZCHEMO-8 may be used to mobilise progenitor cells from  
 XX the marrow into the peripheral blood for transplants. ZCHEMO-8  
 XX polypeptides could be used to further define the role of chemokines in  
 XX mediating suppression of HIV replication in CD4+ T-cells and limiting  
 XX progression of HIV infection to AIDS. Use may be made of ZCHEMO-8  
 XX polypeptides during chemotherapy or radiation therapy, to protect  
 XX haematopoietic cells. ZCHEMO-8 antagonists may have a beneficial  
 XX therapeutic effect in diseases where the inhibition of activation of  
 XX certain macrophages, neutrophils, basophils, B lymphocytes and/or T cells  
 XX may be effective. Such diseases include autoimmune diseases e.g. multiple  
 XX sclerosis, insulin-dependent diabetes and systemic lupus erythematosus,  
 XX rheumatoid arthritis, allergies, asthma or arteriosclerosis. Also benefit  
 XX may be derived from using ZCHEMO-8 antagonists for chronic inflammatory  
 XX and infective diseases. Antagonists may be used to dampen or inactivate  
 XX ZCHEMO-8 during activated immune response. The products can also be used  
 XX for detection, diagnosis, drug screening or gene therapy.  
 XX Sequence 282 BP; 48 A; 29 C; 37 G; 38 T; 130 other;

Query Match 3.1%; Score 34.2; DB 20; Length 282;  
 Best Local Similarity 27.7%; Pred. No. 0.26;  
 Matches 57; Conservative 40; Mismatches 109; Indels 0; Gaps 0;  
 QY 15 actgtttccattctgacagtggttgcacctctctacagcgtgtgtcaggatgc 74  
 Db 281 ARYGTGTTGGNGTGTNARNARNNSWDATRTAYTTTGNCCCAATTTTNCCKNGGTGN 222  
 QY 75 ttagaccattctctgtttctctcagcagtcacagagagtggtctactgtca 134  
 Db 221 GTRCANACTYTTTNCCKYTTNGTNGTRADATNACNGCKYTGNSWRCANSWRTN 162  
 QY 135 gatgttagctgtctctccactgtacacttcaagctgtctctgtgtgaggaaccagaag 194  
 Db 161 SWNGTRAAATCTRANSNCKNACCCANGTCCANGNARNGGYTTTGTGNSWRTAYTGRAAR 102  
 QY 195 ggcctgtccctacttacttactgggcc 220  
 Db 101 CARCANGTGTNNSWDATRTCNWNCC 76

RESULT 7

AL0207  
 ID AL0207 standard; CDNA; 307 BP.  
 XX AC AL0207;  
 XX DT 03-JUL-2000 (first entry)  
 XX DE Rat liver toxicological response marker, SEQ ID NO:100.  
 XX KW Toxicological response marker; rat; liver; expression pattern;  
 KW toxicity screening; toxic compound; polycyclic aromatic hydrocarbon;  
 KW PAH; benzo(a)pyrene; clofibrate; acetaminophen; ss.  
 XX OS Rattus norvegicus.  
 XX PN WO200012760-A2.  
 XX PD 09-MAR-2000.  
 XX PF 27-AUG-1999; 99WO-US19768.  
 XX PR 28-AUG-1998; 98US-0141825.  
 PR 13-OCT-1998; 98US-0172108.  
 PR 13-OCT-1998; 98US-0172711.  
 XX (INCY-) INCYTE PHARM INC.  
 XX Cunningham MJ; Zweiger GB, Panzer SR, Seilhamer JJ;  
 WPI; 2000-237888/20.

Isolated and purified nucleic acid molecules used as toxicological  
 response markers for detecting and diagnosing a potential toxicological  
 response in a mammalian subject to a test compound or molecule -  
 Claim 6; Page 70; 76pp; English.

Sequences AL0108-AL0224 represent rat liver toxicological response  
 markers. These were identified by their pattern of at least twofold  
 upregulation or downregulation of expression in rat liver treated with  
 a toxic compound (e.g., clofibrate, acetaminophen or polycyclic  
 aromatic hydrocarbons (PAHs) such as benzo(a)pyrene). Fluorescently  
 labelled rat liver mRNA was contacted with a microarray comprising a  
 library of rat cDNA molecules. Two-fold or larger changes in hybridisation  
 were only observed between the sample mRNA and sequences AL0108-AL0224.  
 In particular, sequences AL0110, AL0116, AL0117, AL0120, AL0126, AL0133,  
 AL0138, AL0140, AL0142-AL0144, AL0146, AL0149, AL0164, AL0174, AL0185,  
 AL0188, AL0189, AL0201 and AL0205 were all upregulated in samples treated  
 with known toxic compounds relative to untreated samples, while sequences  
 AL0150, AL0156, AL0157, AL0159-AL0163, AL0166, AL0170, AL0175,  
 AL0178, AL0181, AL0192, AL0194, AL0197, AL0202, AL0209, AL0210, AL0212  
 and AL0222 were all downregulated. Expression of these sequences is  
 therefore modulated in liver during a metabolic response to a toxic  
 compound. The markers may be used as probes to determine the toxicity of  
 a test compound. A tissue sample from an animal treated with the test  
 compound is obtained, labelled (e.g., with a fluorophore) and then  
 contacted with a microarray comprising the markers. The expression  
 pattern of the markers may then be compared with the marker expression  
 pattern in untreated control samples, and the toxicity of the test  
 compound determined. The tissue sample is preferably selected from liver,  
 kidney, brain, spleen, pancreas and lung. The nucleic acid molecules of  
 methods of the invention may also be used for screening libraries of  
 molecules for specific binding affinity, and for the fine-tuning of  
 treatment regimens which use drugs with toxic side-effects such that the  
 side-effects are minimised without compromising the efficacy of the drug.

Sequence 307 BP; 86 A; 55 C; 81 G; 79 T; 6 other;  
 Query Match 3.1%; Score 34.2; DB 21; Length 307;





Mon Nov 6 10:14:44 2000

are also used for treating conditions associated with abnormal physiology or development, including inflammatory conditions. The polypeptide cytokine should mediate cytokine synthesis and proliferation in cells. IL-B30 is useful for drug screening to identify compounds having binding affinity to IL-B30.

Sequence 570 BF; 121 A; 168 C; 157 G; 124 T; 0 other;  
 Query Match 3.0%; Score 32.8; DB 20; Length 570;  
 Best Local Similarity 52.1%; Pred. No. 1.1;  
 Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
 QY 629 agaaactggaagtgcaggtccagaggaattttgcctttgtgtcctgagtcaccag 688  
 DB 482 AGAAGGAGACCTGCCATGGCTGGCTGGACTGAGGCTTGGAACTCTGCTGAGTCTCCAG 423  
 QY 689 gcaggtcaactggagcagaaaaattgtttccctcgttcagggcctgaagtgcacc 748  
 DB 422 TGGTGACCTCAGGCTGAGGAGTTGGCTGAGGCCAGTAGGGAGCATGAAGCTGCGCC 363  
 QY 749 tcaggttggtttcagctg 768  
 DB 362 ACAGGGCTATCAGGAGCAG 343

RESULT 11

AL0594/c  
 ID AL0594 standard; DNA; 10732 BP.

XX AC AL0594;

XX DT 29-JUN-2000 (first entry)

XX DE Gene encoding a subunit of cellulose synthase.

XX KW Cellulose synthase; cellulose production; increase yield; ds.

XX OS Vigna angularis.

XX PN JP2000060568-A.

XX PD 29-FEB-2000.

XX PF 26-AUG-1998; 98JP-0239998.

XX PR 26-AUG-1998; 98JP-0239998.

XX PA (MIZUO) MIZUNO K.

XX PA (OJIP) OJI PAPER CO.

XX DR WPI; 2000-342371/30.

XX DR P-PSDB; Y85179.

XX PT A gene encoding a cellulose synthetic equipment - for the improvement in the amount of cellulose synthesised in a plant body  
 XX PS Claim 2; Page 14-21; 32pp; Japanese.  
 XX CC This sequence represents a gene encoding a subunit of the cellulose synthase complex of Vigna angularis. The invention relates to subunits of cellulose synthetic equipment, that can be used to increase the amount of cellulose synthesised by a plant. The proteins and genes encoding them can also be used to improve the properties of the cellulose being produced by a plant.  
 XX CC Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;

Query Match 3.0%; Score 32.4; DB 21; Length 10732;  
 Best Local Similarity 16.5%; Pred. No. 8.6;  
 Matches 71; Conservative 167; Mismatches 189; Indels 3; Gaps 2;

Query Match 3.0%; Score 32.8; DB 20; Length 570;  
 Best Local Similarity 52.1%; Pred. No. 1.1;  
 Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 629 agaaactggaagtgcaggtccagaggaattttgcctttgtgtcctgagtcaccag 688  
 DB 482 AGAAGGAGACCTGCCATGGCTGGCTGGACTGAGGCTTGGAACTCTGCTGAGTCTCCAG 423  
 QY 689 gcaggtcaactggagcagaaaaattgtttccctcgttcagggcctgaagtgcacc 748  
 DB 422 TGGTGACCTCAGGCTGAGGAGTTGGCTGAGGCCAGTAGGGAGCATGAAGCTGCGCC 363  
 QY 749 tcaggttggtttcagctg 768  
 DB 362 ACAGGGCTATCAGGAGCAG 343

RESULT 10

X17786/c  
 ID X17786 standard; cDNA; 570 BP.

XX AC X17786;

XX DT 21-MAY-1999 (first entry)

XX DE Human interleukin-B30 (IL-B30) polypeptide encoding cDNA.

XX KW Cytokine; interleukin-B30; IL-B30; forensic science; cell proliferation; inflammatory condition; drug screening; human; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT 1..570

XX FT /tag= a

XX FT /product= "IL-B30"

XX FT 1..63

XX FT /tag= b

XX FT 64..567

XX FT /tag= c

XX FT sig\_peptide

XX FT mat\_peptide

XX PN WO9905280-A1.

XX PD 04-FEB-1999.

XX PF 24-JUL-1998; 98WO-US15423.

XX PR 25-JUL-1997; 97US-0900905.

XX PA (SCHE) SCHERING CORP.

XX PI Bazan JF;

XX PI WPI; 1999-142935/12.

XX DR P-PSDB; W95002.

XX CC Newly isolated or recombinant polynucleotide encoding mammalian cytokine interleukin-B30 (IL-B30), including fragments - useful for regulating activation, development, differentiation and function of various cell types, and for diagnosing and treating conditions associated with IL-B30  
 XX CC Claim 3; Page 8-9; 83pp; English.  
 XX CC This cDNA encodes a human cytokine interleukin-B30 (IL-B30) polypeptide. Host cells containing a vector comprising the IL-B30 nucleic acid are used for the recombinant production of the protein. The polynucleotides are useful for diagnosis of IL-B30 mediated conditions, and forensic science (e.g. to distinguish rodent from human, or as a marker to distinguish between different cells exhibiting differential expression or modification patterns). The IL-B30 (including fragments), together with antibodies that bind to IL-B30 are useful for teaching purposes. They

QY 256 tctctagagcctgaatgtggcagagtagtctctctggtttctaggtgtctt 315  
 Db 10252 YSCTTSSRSTSRSYDADCSRCNSTYRAYRACRNCSTNCSTDRCSRGYSIDAYATBN 10193  
 QY 316 ccctctgaaggtctagctctccctccatggatgggtgcaggagctgtttgacca 375  
 Db 10192 CYISYNSCTCYTSDCTSCRTBSYDADCSRCNSTYSCYDACSSTYASTDNCNR 10133  
 QY 376 ggtctctcaaatccggtcgctgagacc-gcaggtctcttagctgctgctgca 433  
 Db 10132 SRTTCTCTAKYSYTBSTNCYSYDSRSDNSTAYASTTTTAYDAYDARCTTDSNTRCYRAS 10073  
 QY 434 atctctccgacccagagcaccacaaagtctctctctggccagagatgtggcacaagtgtg 493  
 Db 10072 TYDASRRRCYSYTRASRTSDCTBSYSCYSYDASRVANCDTBCYTTBAKYRACYD 10013  
 QY 494 ggcagaagtgcactctctctgctccctagc-gtctcaggattgcccctacttgggcaa 552  
 Db 10012 AYAKRNCSTIRAYSSTCTYRCRCRNCYSTYSYSTYASTTBTTCYCTBCSRRC 9953  
 QY 553 tccgctctctctccacaggggtttgggagcaggagctgtggccggtatcaggcacaagg 612  
 Db 9952 YSSRYSTNCYSYCCYTSRYSTTNTCTCRCTYSYNSYTBYSYSSYSSRGYSRG 9893  
 QY 613 tttaggcaaccaggttagaactggaagtgtcaggtgtccagaggaatttgcctttgtgt 672  
 Db 9892 SDSRGNCYNTNCYDASTDTSYRCCTYSYSSDSTSDTSTNSTBSCDYTTTBS 9833  
 QY 673 gtcctgagtc 682  
 Db 9832 RSTSDSTSTY 9823

## RESULT 12

X87772 ID X87772 standard; DNA; 3835 BP.  
 XX AC X87772;  
 XX  
 DT 09-NOV-1999 (first entry)  
 XX  
 DE Human mucin gene MUC 5AC 5' regulatory region.  
 XX  
 KW Mucin gene; MUC 5AC; mucomone; human; promoter; inhibitor;  
 KW screening; cystic fibrosis; Pseudomonas aeruginosa; infection;  
 KW tobacco smoke; bronchial pneumonia; chronic bronchitis; asthma; ss.  
 XX  
 OS Homo sapiens.  
 XX

Key Location/Qualifiers  
 FH misc\_signal 1..218  
 FT /\*tag= a  
 FT /note= "mucomone response element of Claim 5"  
 FT /\*tag= b  
 FT /\*tag= c  
 FT /note= "mucomone response element of Claim 5"  
 FT /note= "mucomone response element of Claim 5"  
 FT /note= "mucomone response element of Claim 5"  
 FT /tag= d  
 FT /partial  
 FT /note= "base 3753 represents the transcription  
 start site"  
 FT 3683..3835  
 FT /\*tag= e  
 FT /partial  
 FT /note= "5' end of coding region"

W09941270-A1.

19-AUG-1999.

PF 11-FEB-1999; 99WO-US03024.  
 XX  
 PR 11-FEB-1998; 98US-0074398.  
 XX  
 PA (REGC) UNIV CALIFORNIA.  
 XX  
 PI Basbaum C, Gallup M, Gebremichael A, Gensch E, Li D;  
 XX WPI; 1999-518437/43.  
 XX  
 DR Inhibition of mucin overexpression, useful for treating, e.g. cystic  
 XX fibrosis, chronic bronchitis or asthma  
 PT  
 PS Claim 1; Fig 1A-B; 58pp; English.  
 XX  
 CC This is the nucleotide sequence of the 5' regulatory region of  
 CC the human mucin gene MUC 5AC. It forms part of an isolated 4 kb  
 CC genomic DNA fragment deposited as ATCC 98701. The 5' regulatory  
 CC region includes mucomone response elements that are responsible for  
 CC an increase in transcription of the MUC 5AC gene, and consequent  
 CC increase in mucin production in vivo, in response to certain  
 CC effectors, i.e. mucomones, particularly environmental tobacco smoke  
 CC or Pseudomonas aeruginosa-conditioned medium. These regulatory  
 CC elements, particularly in combination with a detectable reporter  
 CC gene, are used to identify compounds that inhibit the induction of  
 CC MUC 5AC by mucomones. Such inhibitors are used for treatment of  
 CC diseases or conditions associated with inappropriate production of  
 CC mucin, especially respiratory diseases such as cystic fibrosis,  
 CC chronic bronchitis, bronchial pneumonia and asthma.  
 XX  
 SQ Sequence 3835 BP; 834 A; 1301 C; 686 G; 1014 T; 0 other;

Query Match 2.9%; Score 31.6; DB 20; Length 3835;  
 Best Local Similarity 56.9%; Pred. No. 8.5;  
 Matches 58; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
 QY 470 gggcccaaggaggtgggcaagggtggcagagtgaggcaatctctctgctgctgctca 529  
 Db 164 gggcccaaggaggtgggcaagggtggcagagtgaggcaatctctctgctgctgctca 223  
 QY 530 ggattgcttcttctggtggcaatccgctctctctccacag 571  
 Db 224 ggcctgctctgtgtggggctggacagctaccaccactg 265

## RESULT 13

X20325/c ID X20325 standard; DNA; 1295 BP.  
 XX AC X20325;  
 XX  
 DT 04-MAY-1999 (first entry)  
 XX  
 DE Borrelia burgdorferi polynucleotide sequence #78.  
 XX  
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.  
 XX  
 OS Borrelia burgdorferi.  
 XX  
 PN W09858943-A1.  
 XX  
 PD 30-DEC-1998.  
 XX  
 PF 18-JUN-1998; 98WO-US12764.  
 XX  
 PR 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX



DT 01-FEB-2000 (first entry)  
XX SGRF coding sequence.  
DE SGRF; human; Interleukin-6 G-CSF related factor; cell proliferation;  
KW immune system; haematopoietic system; therapy; ss.  
XX Homo sapiens.  
XX WO954357-A1.  
PN 28-OCT-1999.  
XX 14-APR-1999; 99WO-JP01997.  
XX 14-APR-1998; 98JP-0121805.  
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX Hirata Y;  
PI WPI; 2000-013230/01.  
DR P-PSDB; Y54606.  
XX Novel cytokine-like protein, with activity of supporting proliferation  
PT of myeloid cells, useful in treating abnormality of cell proliferation  
PT in immune and haematopoiesis systems  
XX Claim 2; Fig 1; 69pp; Japanese.  
XX This sequence encodes the Interleukin-6 G-CSF related factor (SGRF)  
CC protein of the invention. The protein is a member of the IL-6/G-CSF/MSF  
CC family. The protein can be used in drugs for treating diseases due to  
CC abnormality of cell proliferation in the immune system and haematopoietic  
CC system.  
XX Sequence 1026 BP; 268 A; 249 C; 272 G; 237 T; 0 other;

Query Match 2.9%; Score 31.2; DB 21; Length 1026;  
Best Local Similarity 51.4%; Pred. No. 5, 3;  
Matches 72; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
QY 629 agaaactggaagtgcaggtccacagagaatttgccttctgtctcagtcacaccag 688  
DB 625 AGAAGGAGACGCTGCCATGCTGGCTGGGACTGAGCTTGAATCTGCTCTCCAG 566  
QY 689 gcaggtcacttgagcagaaaaattggtttccctcgtctcaggtcaggtcaggtgacc 748  
DB 565 TGGTGACCTCAGGCTGCAGGAGTTGGCTGAGGCCCACTAGGGAGGCATGAAGCTGGCCC 506  
QY 749 tcagggttggtttcagctg 768  
DB 505 ACAGGGCTATCAGGAGCAG 486

Search completed: November 4, 2000, 00:32:12  
Job time: 21602 sec



Query Match 3.7%; Score 40.8; DB 1; Length 7218;  
Best Local Similarity 6.6%; Pred. No. 0.0043;  
Matches 27; Conservative 204; Mismatches 181; Indels 0; Gaps 0;

QY 1 ttgtctctgtgttacttctttccatttgcagtggtttgacctctctacgctg 60  
DB 1088 YY 1147  
QY 61 ttgtcagagtgctgtacacatttctctctttcttcttcagcaggttaaggaacaga 120  
DB 1148 YY 1207  
QY 121 ttgttactgtcagtggtgtctgtctctcactgactttcaagctctctgtgt 180  
DB 1208 YY 1267  
QY 181 gcaggaaccagagggcctgtccctactctactgtgcccctacacaggggcctaga 240  
DB 1268 YY 1327  
QY 241 ttgtgtcagtggtttctctctagaccctgaatgtgggcagagagtagtctctctgtt 300  
DB 1328 YY 1387  
QY 301 tctaggtatgtcttccctctgaaggtctagctctctctctccatgggatgggtcag 360  
DB 1388 YY 1447  
QY 361 gtagctgttgacaggtctctcctcaatcccggtgagctgtggaccgaggc 412  
DB 1448 CTATCTCTTAATACTTGCATAGATAGTAAATTACAGTGATGCTACATGC 1499

RESULT 2  
US-08-889-402-4/C  
; Sequence 4, Application US/08889402  
; Patent No. 5811288  
; GENERAL INFORMATION:  
; TITLE OF INVENTION: NOVEL ISOFORM GENE FOR FOCAL ADHESION  
; NUMBER OF SEQUENCES: 7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/889,402  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 178334/1996  
; FILING DATE: 08-JUL-1996  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1818 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CD3  
; LOCATION: 1..1815  
US-08-889-402-6

Query Match 3.1%; Score 33.4; DB 2; Length 1818;  
Best Local Similarity 53.4%; Pred. No. 0.51;  
Matches 70; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 174 tctgtgtcaggaacagagggcctgtctctacttctacttctctacgagggg 233  
DB 1073 TCAGACTGCAGGCTCCCGCAGCATGCTGCCAGCTGGCTCCGGGGCTTCCGGGCCCA 1014  
QY 234 gcctagatgg:gctaggtgtttctctctagagcctgaaatgtggcagagagtagtctcc 293  
DB 1013 GGGGGTGAGCTGCTCCCTGTCTTCCCTGGGCCATGAACATGTGGAGAGGGGGCTGG 954  
QY 294 tctgtgttctcc: 304  
DB 953 TCAGGACTCCT: 943

RESULT 4  
US-09-122-443-1/c  
; Sequence 1, Application US/09122443  
; Patent No. 6060284  
; GENERAL INFORMATION:  
; APPLICANT: Bazar, J. Fernando  
; TITLE OF INVENTION: MAMMALIAN CYTOKINE; RELATED REAGENTS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104

DB 1013 GGGGGTGAGCTGCTCCCTGTCTTCCCTGGGCCATGAACATGTGGAGAGGGGGCTGG 954  
QY 294 tctgtgttctcc: 304  
DB 953 TCAGGACTCCT: 943

RESULT 3  
US-08-889-402-6/c  
; Sequence 6, Application US/08889402  
; Patent No. 5811288  
; GENERAL INFORMATION:  
; TITLE OF INVENTION: NOVEL ISOFORM GENE FOR FOCAL ADHESION  
; NUMBER OF SEQUENCES: 7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/889,402  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 178334/1996  
; FILING DATE: 08-JUL-1996  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1818 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CD3  
; LOCATION: 1..1815  
US-08-889-402-6

Query Match 3.1%; Score 33.4; DB 2; Length 1818;  
Best Local Similarity 53.4%; Pred. No. 0.51;  
Matches 70; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 174 tctgtgtcaggaacagagggcctgtctctacttctacttctctacgagggg 233  
DB 1073 TCAGACTGCAGGCTCCCGCAGCATGCTGCCAGCTGGCTCCGGGGCTTCCGGGCCCA 1014  
QY 234 gcctagatgg:gctaggtgtttctctctagagcctgaaatgtggcagagagtagtctcc 293  
DB 1013 GGGGGTGAGCTGCTCCCTGTCTTCCCTGGGCCATGAACATGTGGAGAGGGGGCTGG 954  
QY 294 tctgtgttctcc: 304  
DB 953 TCAGGACTCCT: 943

RESULT 4  
US-09-122-443-1/c  
; Sequence 1, Application US/09122443  
; Patent No. 6060284  
; GENERAL INFORMATION:  
; APPLICANT: Bazar, J. Fernando  
; TITLE OF INVENTION: MAMMALIAN CYTOKINE; RELATED REAGENTS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/122,443  
FILING DATE: 24-JUL-1998  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/053,765  
FILING DATE: 25-JUL-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0758K1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650)852-9196  
TELEFAX: (650)496-1200  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 570 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..567  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 64..567  
US-09-122-443-1

Query Match 3.0%; Score 32.8; DB 5; Length 570;  
Best Local Similarity 52.1%; Pred. No. 0.4;  
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
QY 629 agaaactggaaggtcagggtccagaggaatttgccttctgtctcgtcagtcacacag 688  
|||||  
DB 482 ACAAGGAGACGCTGCCATGCTGGCTGGAGCTTGGAACTGGAGTCTGAGTCTCCAG 423  
QY 689 cgaggtcactgagcagaaaaattgggtttccctcgtgtctcagcctgagttgacc 748  
|||||  
DB 422 TGGTGACCTCAGGCTGCAGGAGTTGGCTAGGCCCCAGTAGGAGGCGATGAGCTGCC 363  
QY 749 tcagggttggtttcagctg 768  
|||||  
DB 362 ACAGGCTATCAGGGAGCAG 343

RESULT 5  
US-08-440-856A-9  
Sequence 9, Application US/08440856A  
Patent No. 5750873  
GENERAL INFORMATION:  
APPLICANT: DELLAPORTA, STEPHEN L.  
TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING  
TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 PENNSYLVANIA AVE. N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/440,856A  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: MILLMAN, ROBERT A.  
REGISTRATION NUMBER: 36,217  
REFERENCE/DOCKET NUMBER: 05463-20001.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1517  
TELEFAX: (202) 887-0763  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1288 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-440-856A-9

Query Match 2.9%; Score 31.6; DB 2; Length 1288;  
Best Local Similarity 35.5%; Pred. No. 1.6;  
Matches 61; Conservative 31; Mismatches 80; Indels 0; Gaps 0;  
QY 139 ttagctgttctcactgactttcaagctgtctctgtgtcaggaacacagaagggcc 198  
|||||  
DB 121 TCYMGCTATCCTGCTGCGTGCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 180  
QY 199 tgcctctacttctactgggccccctacgcacagggggcctagatggtgtgtgtttcc 258  
|||||  
DB 181 TGRKCCSTCKWCCACGSGCTGGGAGGCAATGSCCAGGCGYSCCCACWCCSATGC 240  
QY 259 tctagagcctgaaatgtgggcagagagtagtctctctctctctctctctctctctat 310  
|||||  
DB 241 CCAAGAGGCTGGAMGGGAAGTGGCCATTGTSACSGGCGGCGMGSGGGAT 292

RESULT 6  
US-07-945-283-1  
Sequence 1, Application US/07945283  
Patent No. 5352596  
GENERAL INFORMATION:  
APPLICANT: Cheung, Andrew K.  
APPLICANT: Wesley, Ronald D.  
TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
TITLE OF INVENTION: Involving The EPO and LIT Genes  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis P. Ribando  
STREET: 1815 No. 5352596th University Street  
CITY: Peoria  
STATE: IL  
COUNTRY: USA  
ZIP: 61604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/945,283  
FILING DATE: 19920911  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ribando, Curtis P  
REGISTRATION NUMBER: 27976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 309-685-4011 ext. 513  
TELEFAX: 309-685-4128  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 8438 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Pseudorabies virus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 622..6495  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1099, "g")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1267, "t")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1381, "c")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1566, "c")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(7010, "g")  
US-07-945-283-1

Query Match 2.9%; Score 31.4; DB 1; Length 8438;  
Best Local Similarity 47.3%; Pred. No. 5.7;  
Matches 95; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
QY 216 ggccccctacgcacaggggcccctagatggtgctaggtgttttctcttagagcctgaatgt 275  
Db 7357 GCTACTTTGCTATGTCGGCCCGGAGGGCCATCTTGGCCCTCGAGCGCAATGGGAT 7416  
QY 276 gggcgagagtagtctctctggttcttaggtatgtcttccctctgaaggtctagctc 335  
Db 7417 TTCTCTCCCTACTTCTCTCGGGTCTACTTTGCTATGTCGGCCCGGCGGCCATCTC 7476  
QY 336 tcccttcctatggatagggcgacggagcgtgttgaccaggctctcctaaccgggtg 395  
Db 7477 GCTCGCCGGCCCAATGGCGCGGAGGGGTCTCCCGCGGCTCTGCTATTTGCTATGCC 7536  
QY 396 cagctggaccgcaggctctc 416  
Db 7537 CGGCGCGCTCTCGGCCATCT 7557

RESULT 7  
US-07-879-617A-7/c  
; Sequence 7, Application US/07879617A  
; Patent No. 5580775  
; GENERAL INFORMATION:  
; APPLICANT: Fremeneau Jr., Robert T.  
; APPLICANT: Caron, Marc G.  
; APPLICANT: Blakely, Randy D.  
; TITLE OF INVENTION: A High Affinity L-Proline Transporter  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick & Cody  
; STREET: 1100 Peachtree Street, Suite 2800  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.  
; ZIP: 30309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/879,617A  
FILING DATE: 19920501  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: EMU109  
TELEPHONE: 404-815-6508  
TELEFAX: 404-815-6555  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2728 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus  
TISSUE TYPE: Brain  
IMMEDIATE SOURCE:  
LIBRARY: rat forebrain cDNA library  
CLONE: rTB2-2-20  
US-07-879-617A-7

Query Match 2.9%; Score 31.2; DB 1; Length 2728;  
Best Local Similarity 48.8%; Pred. No. 3.4;  
Matches 84; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
QY 417 gtacgttgcctggcgaattcttcacccagagagagcccccagttctcttgggccc 476  
Db 2493 GTCGCTGCCATCTTATCTATATCCAGGCGCTTGGGCTCTCCAGGTTCTGCTACCTGA 2434  
QY 477 ggaatggggcagaagggtggcagaagtgcaattctctccgctcctagcctcagattgc 536  
Db 2433 ACCTGGGTTTAAAGTGGCTGTATGGAATGTATCTGTCAAGTGTCTGAACAAGA 2374  
QY 537 cctcaattctggcaatccgctctctcttccacaggggttggagcaggag 588  
Db 2373 TGTGATATATGTCCTGCTCTCCCAAGGAGGAGTTGACTGGCTGGAG 2322

RESULT 8  
US-08-753-985-7/c  
; Sequence 7, Application US/08753985  
; Patent No. 5759788  
; GENERAL INFORMATION:  
; APPLICANT: Fremeneau Jr., Robert T.  
; APPLICANT: Caron, Marc G.  
; APPLICANT: Blakely, Randy D.  
; TITLE OF INVENTION: A High Affinity L-Proline Transporter  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick & Cody  
; STREET: 1100 Peachtree Street, Suite 2800  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: U.S.  
; ZIP: 30309  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/753,985  
; FILING DATE: 03-DEC-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:



APPLICATION NUMBER: US 07/879617  
FILING DATE: 01-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: EMU109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-815-6508  
TELEFAX: 404-815-6555  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2728 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus  
TISSUE TYPE: Brain  
IMMEDIATE SOURCE:  
LIBRARY: rat forebrain cDNA library  
CLONE: rTB2-2-20  
US-08-733-985-7

Query Match 2.9%; Score 31.2; DB 2; Length 2728;  
Best Local Similarity 48.8%; Pred. No. 3.4;  
Matches 84; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
QY 417 gtatgttgcctgtcgaattctccgcacccagagaccacaaagttctcttgggcca 476  
DB 2493 GTCGCTGCCATTCTATCTATATATCCAGGCGCTTGGGCTCTCAGGTTCTCGTACCCCTGA 2434  
QY 477 ggaatgggcaagtgggcagaagtggaattctctcctgcctagcgtctcaggattgc 536  
DB 2433 ACCTGGGGTTTAAGTTGGCTGTATGGAATGATCTGTCAAGTCTCIGACAAAGA 2374  
QY 537 cctcactcttgggcaatcgcgtctctcttccacagggttggagcagggag 588  
DB 2373 TGTCAGATATGCTGTGCTGCCCAAGGAAAGGAGCTTGACTGCTGGAG 2322

RESULT 9  
PCT-US94-888-497-33  
Sequence 33, Application US/08888497  
Patent No. 5972677  
GENERAL INFORMATION:  
APPLICANT: Tischfield, Jay A.  
APPLICANT: Seilhamer, Jeffrey J.  
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide  
TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences  
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide  
TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
ADDRESSEE: Russell PA  
STREET: 200 East Broward Boulevard  
CITY: Fort Lauderdale  
STATE: FL  
COUNTRY: USA  
ZIP: 33301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/0888,497  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,405  
FILING DATE:  
APPLICATION NUMBER: US 08/097,354  
FILING DATE: 26-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Manso, Peter J.  
REGISTRATION NUMBER: 32,264  
REFERENCE/DOCKET NUMBER: IN21044-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 305-527-2498  
TELEFAX: 305-764-4996  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15328 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-888-497-33

Query Match 2.8%; Score 31; DB 4; Length 15328;  
Best Local Similarity 47.6%; Pred. No. 11;  
Matches 91; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 97 ctttcagccagttacaggaagtgcttctactgtcagatgtgttagctgttctctgtcca 156  
DB 4206 CTCACAGTAGTGTCTCAGTAAAGTGTTAAATGAATGAATGGGCTAGTTGTCTCT 4265  
QY 157 ctgacttcacagctgtctctgtgaggaaccaggaagggcctgtccctacttctactgg 216  
DB 4266 GGGTCATCATCTTCCAGCTGCCTAAAGTTGGGAAATTTGGCAATCTCAGTCCC 4325  
QY 217 gcccctacgcagggggccctagatggtctaggtgttttctctctagagcctgaaatgg 276  
DB 4326 TCCCTACAAAGGCGAGCAATGATTGTACTTTATATAGTTTCTAGTAGTAATGAGATAGC 4385  
QY 277 ggcagagagta 287  
DB 4386 AACAGATACIA 4396

RESULT 10  
PCT-US94-07926-33  
Sequence 33, Application PC/TUS9407926  
GENERAL INFORMATION:  
APPLICANT: Tischfield, Jay A.  
APPLICANT: Seilhamer, Jeffrey J.  
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide  
TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences  
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide  
TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
ADDRESSEE: Russell PA  
STREET: 200 East Broward Boulevard  
CITY: Fort Lauderdale  
STATE: FL  
COUNTRY: USA  
ZIP: 33301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07926  
FILING DATE: 15-JUL-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/097,354

FILING DATE: 26-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Manso, Peter J.  
REGISTRATION NUMBER: 32,264  
REFERENCE/DOCKET NUMBER: IN21044-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 305-527-2498  
TELEFAX: 305-764-4996  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15328 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
PCT-US94-07926-33

Query Match 2.8%; Score 31; DB 6; Length 15328;  
Best Local Similarity 47.6%; Pred. No. 11;  
Matches 91; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 97 cttcagccaggtacaggaacagaggtttctactgtcagatgtgtagctgttctctgtcca 156  
Db 4206 CTCACAGTAGGTCCTCAGTAAAGAGTGTTAAATGAATGAATGGGCTAGGTTGTGCTC 4265  
QY 157 ctgacttcaagctgtctgtgtgcaggaaacagagagggccctgtccctacttctactgg 216  
Db 4266 GGGTCTATCATCTCCAGCTGCTAAGTTGGGAATGGGCTCTGTGAAATCTCAGTCCC 4325  
QY 217 gccctacagcagaggggctcagatggtgctaggtgtttctctctagagccctgaaatgtg 276  
Db 4326 TCCCTACAAAAGGCGACGAATGATTGTTACTTATAGTTTCTAGTAGTAGTAATGAGATAGC 4385  
QY 277 ggcagagagta 287  
Db 4386 AACAGATACTA 4396

RESULT 11  
US-08-356-060A-6/c  
Sequence 6, Application US/08356060A  
Patent No. 5844079  
GENERAL INFORMATION:  
APPLICANT: Ingham, Phillip W.  
APPLICANT: McMahon, Andrew P.  
APPLICANT: Tabin, Clifford J.  
TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing  
TITLE OF INVENTION: Proteins and Uses Related Thereto  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,060A  
FILING DATE: 14-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/176,427  
FILING DATE: 30-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: HMI-006CP

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1425 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1425  
US-08-356-060A-6

Query Match 2.8%; Score 30.6; DB 3; Length 1425;  
Best Local Similarity 48.6%; Pred. No. 3.6;  
Matches 84; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 343 catgggatagtggtgcaggagagctgtttgaccaggtctctcaaatccgggtgagctg 402  
Db 480 CATGCCGTACTTCTGCGGTGCGGTCAGACGTGGTGATGTCACATGCGCGGCCCTCGTA 421  
QY 403 gacgcagggctcctgtagcttgctctgcaatcttccgcacccagagagaccacagtt 462  
Db 420 GTCAGAGACTCTCTGAGTGGGCCATCTTCGTCCCGACCCCTGGTCACCGCAGTTT 361  
QY 463 tctcttggggccaagatgtgggcaagtggtggcagaagtggcaatctctct 515  
Db 360 CACTCTCTGGCCTACTGTTTCATCACCAGATGCCCAAGCGTTCACTTGTCTCT 308

RESULT 12  
US-08-748-591-5/c  
Sequence 5, Application US/08748591  
Patent No. 5759811  
GENERAL INFORMATION:  
APPLICANT: Epstein, Ervin  
APPLICANT: Hu, Zhilan  
APPLICANT: Bonifas, Jeanette  
TITLE OF INVENTION: Mutant Human Hedgehog Gene  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish and Richardson  
STREET: 2200 Sand Hill Road  
CITY: Menlo Park  
STATE: CA  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/748,591  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sherwood, Pamela J  
REGISTRATION NUMBER: 36,677  
REFERENCE/DOCKET NUMBER: 06510/067001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 322-5070  
TELEFAX: (415) 854-0875  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1576 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna

US-08-748-591-5

Query Match 2.8%; Score 30.6; DB 2; Length 1576;  
Best Local Similarity 48.6%; Pred. No. 3.9;  
Matches 84; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 343 catggatagggcaggagctgttgaccaggtctctcaaatccgggtgagctcg 402  
DB 631 CARGCGGTACTGTCGCGGTGCGGTGACAGCTGTGATGTCACCTGCGCGGCCCTCCTA 572  
QY 403 gaccgcaggtcgttagcttgcctgctgcaatctcccgaccagagggcaccacgaatt 462  
DB 571 GTGCAGAGACTCCTCTGAGTGTAGCCATCTTCGTCGCCAGCCCTCGGTCAACCGCAGTT 512  
QY 463 tctcttggccaaagatgtggtggaaggtgggcaagtggtgcaatctctctct 515  
DB 511 CACTCTGCGCCACTGTTTCATCACCAGAGATGCCCAAGCGTTCAACTTGTCT 459

RESULT 13

US-08-748-591-10/c  
; Sequence 10, Application US/08748591  
; Patent No. 5759811  
; GENERAL INFORMATION:

APPLICANT: Epstein, Ervin  
APPLICANT: Hu, Zhilan  
APPLICANT: Bonifas, Jeanette  
TITLE OF INVENTION: Mutant Human Hedgehog Gene  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish and Richardson  
STREET: 2200 Sand Hill Road  
CITY: Menlo Park  
STATE: CA  
COUNTRY: USA  
ZIP: 94025

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/748,591  
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Sherwood, Pamela J  
REGISTRATION NUMBER: 36,677  
REFERENCE/DOCKET NUMBER: 06510/067001  
TELEPHONE: (415) 322-5070  
TELEFAX: (415) 854-0875  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1576 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-748-591-10

Query Match 2.8%; Score 30.6; DB 2; Length 1576;  
Best Local Similarity 48.6%; Pred. No. 3.9;  
Matches 84; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 343 catggatagggcaggagctgttgaccaggtctctcaaatccgggtgagctcg 402  
DB 631 CATGCCGTACTTGTGCGGTGCGGTGACAGCTGTGATGTCACCTGCGCGGCCCTCGTA 572  
QY 403 gaccgcaggtcctgtgtagctgctgctgcaatctctcccgaccagagggcaccacgaatt 462

DB 571 GTGCAGAGACTCCTCTGAGTGTGGCCATCTTCGCCAGCCCTCGTCCACCGCAGTTT 512

QY 463 tctcttggccaaagatgtggtggaaggtgggcaagtggtgcaatctctctct 515  
DB 511 CACTCTGCGCCACTGTTTCATCACCAGAGATGCCCAAGCGTTCAACTTGTCT 459

RESULT 14

US-08-580-680-2  
; Sequence 2, Application US/08580680  
; Patent No. 5766866  
; GENERAL INFORMATION:

APPLICANT: Center, David M.  
APPLICANT: Cruikshank, William W.  
APPLICANT: Kornfeld, Hardy  
TITLE OF INVENTION: LYMPHOCYTE CHEMOATTRACTANT FACTOR AND  
TITLE OF INVENTION: USES THEREOF  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530-0299

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/580,680  
FILING DATE: 29-DEC-1995  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/480,156  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Frank S. Digiglio  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 98742Y  
TELEPHONE: (516) 742-4343  
TELEFAX: (516) 742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2150 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-580-680-2

Query Match 2.8%; Score 30.6; DB 2; Length 2150;  
Best Local Similarity 58.1%; Pred. No. 4.7;  
Matches 54; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 235 cctagatggtgctaggtgttttctctcttagagcctgaaatgtgggcagagtagtctct 294  
DB 1531 CCCAACAGTGCATCTTTCTCATGAATCTGCAATGTGGGCGAGAGATTGAATGGG 1590  
QY 295 ctggttctcctaggtatgtcttccctctgaagg 327  
DB 1591 CAGCTCATCTCTGCCACTTGGCATCAGCTGG 1623

RESULT 15

US-08-480-156A-2  
; Sequence 2, Application US/08480156A  
; Patent No. 5807549  
; GENERAL INFORMATION:

Search completed: November 4, 2000, 00:13:27  
Job time: 22211 sec

Mon Nov 6 10:14:46 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 3, 2000, 22:08:03 ; Search time 3273.53 Seconds  
(without alignments)  
2058.713 Million cell updates/sec

Title: US-09-101-423A-6  
Perfect score: 1090  
Sequence: 1 ttgtctctgtgttacttgt.....cctagcacagtgcattccaca 1090

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*

44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estom.\*  
68: em\_estov1.\*  
69: em\_estov2.\*  
70: em\_estpl1.\*  
71: em\_estpl2.\*  
72: em\_estpl3.\*  
73: em\_estpl4.\*  
74: em\_estpl5.\*  
75: em\_estro1.\*  
76: em\_estro2.\*  
77: em\_estro3.\*  
78: em\_estro4.\*  
79: em\_estro5.\*  
80: em\_estro6.\*  
81: em\_estro7.\*  
82: em\_estro8.\*  
83: em\_estro9.\*  
84: em\_estro10.\*  
85: em\_estro11.\*  
86: em\_estro12.\*  
87: em\_estro13.\*  
88: gb\_gss1.\*  
89: gb\_gss2.\*  
90: gb\_gss3.\*  
91: gb\_gss4.\*  
92: em\_gss1.\*  
93: em\_gss2.\*  
94: em\_gss3.\*  
95: em\_gss4.\*  
96: gb\_gss5.\*  
97: gb\_gss6.\*  
98: gb\_gss7.\*  
99: gb\_gss8.\*  
100: gb\_gss9.\*  
101: em\_gss5.\*  
102: em\_gss6.\*  
103: em\_gss7.\*  
104: em\_gss8.\*  
105: em\_gss9.\*  
106: em\_gss10.\*  
107: em\_gss11.\*  
108: gb\_gss10.\*  
109: gb\_gss11.\*  
110: em\_gss12.\*  
111: gb\_gss12.\*  
112: gb\_gss13.\*  
113: gb\_gss14.\*  
114: gb\_gss15.\*  
115: gb\_gss16.\*  
116: gb\_gss17.\*

117: gb\_gss18: \*  
118: gb\_gss19: \*  
119: em\_gss13: \*  
120: gb\_gss20: \*  
121: gb\_gss21: \*  
122: gb\_gss22: \*  
123: gb\_gss23: \*  
124: gb\_gss24: \*  
125: em\_gss14: \*  
126: em\_gss15: \*  
127: em\_gss16: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	154.4	14.2	576	33	BE101139	UI-R-BJ1-
C 2	154	14.1	305	33	BE113924	UI-R-BJ1-
C 3	153.4	14.0	305	33	BE113975	UI-R-BJ1-
4	144.8	13.3	493	36	C07078	Rat
5	142.4	13.1	368	37	H33201	EST108975 R
6	124.6	11.4	353	37	H32857	EST108352 R
7	118.4	10.9	505	25	AW921512	R
8	114.8	10.5	493	33	AW921512	EST352816
9	102.2	9.4	254	11	AI237080	EST233642
10	102	9.4	232	12	AI555421	UI-R-C2P-
C 11	100.2	9.2	616	25	AI715774	UI-R-YO-a
12	99.2	9.1	373	33	AW920598	EST351902
13	93.2	8.6	460	25	BE109582	UI-R-BJ1-
C 14	91.8	8.4	507	8	AW921250	EST352554
15	83.2	7.6	457	9	AI137948	UI-R-CO-h
C 16	76.8	7.0	182	37	AI176946	EST220552
17	75.6	6.9	337	8	H32981	EST108559 R
C 18	75.2	6.9	608	10	AI029941	UI-R-CO-1
C 19	73.4	6.7	442	33	AI407793	EST236083
C 20	71.2	6.5	258	37	BE096067	UI-R-C4-a
C 21	70.8	6.5	319	7	H32274	EST107204 R
C 22	70.4	6.5	329	11	AF209968	AF209968
C 23	68.2	6.3	511	33	AI502036	UI-R-EO-d
C 24	67.6	6.2	588	36	BE112904	UI-R-BJ1-
C 25	66	6.1	292	8	C07053	C07053 Rat
C 26	62.2	5.7	301	37	AI013526	EST208201
C 27	60.8	5.6	588	7	H33584	EST109725 R
C 28	58.6	5.4	226	33	AI009160	EST203611
C 29	56.8	5.2	516	8	BE112348	UI-R-BJ1-
C 30	56.8	5.2	549	25	AW915783	EST206851
C 31	56.4	5.2	302	22	AW529404	UI-R-BJ1-
C 32	55.6	5.1	234	22	AW534811	UI-R-BJ1-
C 33	54.8	5.0	568	113	AW95864	RPCI-23-3
C 34	53.4	4.9	352	8	AI072604	UI-R-C2-n
C 35	51.4	4.7	422	114	AZ029250	RPCI-23-3
C 36	51.4	4.7	474	6	AA818642	UI-R-AO-a
C 37	50.2	4.6	407	9	AI232935	EST29623
C 38	47.8	4.4	492	33	BE099466	UI-R-BJ1-
C 39	47.6	4.4	603	114	AZ041297	RPCI-23-3
C 40	47.4	4.3	341	22	AW532384	UI-R-BJ1-
C 41	47.4	4.3	462	37	H31593	EST105773 R
C 42	46.8	4.3	613	113	AZ015508	RPCI-23-3
C 43	44.4	4.1	289	33	BE099429	UI-R-BJ1-
C 44	44	4.0	649	117	AZ252318	RPCI-23-4
45	43.6	4.0	236	36	C06741	Rat

## ALIGNMENTS

RESULT 1  
BE101139

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BE101139: 576 bp mRNA EST 13-JUN-2000  
UI-R-BJ1: auh-e-09-0-UI.s1 UI-R-BJ1 Rattus norvegicus CDNA clone  
UI-R-BJ1: auh-e-09-0-UI 3', mRNA sequence.

BE101139  
BE101139.1 GI:8493048  
EST

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

1 (bases 1 to 576)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene  
discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dr track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dr track served to identify it as a clone from the  
normalized ventricle at 13 dpc library cDNA Library Preparation:  
M.B. Soares Lab Clone distribution: clones will be available  
through Research Genetics (www.resgen.com) The following repetitive  
elements were found in this cDNA sequence: 5-575, >LI\_RN#LINE/LI

Seq primer: M13 Forward

POLYA=Yes.

Location/Qualifiers

1..576

source

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-BJ1-auh-e-09-0-UI"

/clone\_lib="UI-R-BJ1"

/lab\_host="DH10B (Life Technologies)"

/note="vector: pMT3D-pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-BJ1

library is a subtracted library derived from the following

tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV

canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,

AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.

For a detailed description of the library from which this

clone was derived, please visit our web site at

ratest.eng.uiowa.edu. The subtraction has been previously

described in (Bonaldo, Lennon and Soares, Genome Research

6:791-806, 1996)

TAG\_LIB=UI-R-BJ1

TAG\_TISSUE=ventricle at 13 dpc

TAG\_SEQ=CAGCGA"

BASE COUNT 87 a 140 c 167 g 182 t

ORIGIN

Query Match 14.2%; Score 154.4; DB 33; Length 576;

Best Local Similarity 65.0%; Pred. No. 5.4e-35;

Matches 260; Conservative 0; Mismatches 136; Indels 4; Gaps 2;

QY 145 tttctctgctcaactgactttcaagctctctctgtgtcaggaaccagagggcctgtccc 204

Db 174 TGGTCATGGGTGTGTACTTACAGGCGGTCACGACCAACAGAGATCTCGGCC 233

QY 205 tactctactggtggccctacgacaggggcctagatggtgctgttctctctaga 264

Db 234 TCCCTTTCTGAGACTTCTGTGACACAGGGTCCAGATGGCGTTTGGTGTTCCTCTGGA 293

```

265 gctgaaatgtggcagagtagtctctctgttctcttagttagttagtcttccctctga 324
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
294 GTCATAGATGCGCAGAGTGCAGTCTCTTCTGTTTCCAGGCGATGCTCGCTCTCTGA 353
Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
325 aggtctagctctctctccatggataggtgagga--gctgttgacaggtccctc 382
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
354 ATGTTAGCTCTCTCTCCAGGAGTGTGTGTCAGAGAACTGTATCTGTCGGTCCCTT 413
Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
383 tcaaatcgggtgcagctctgacccaggtctctctgttagtctgctgctgcaatctcccg 442
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
414 TAGGTTCTGGGGTGTCTCAGACCGACCGGACCTGCTGCTCCCTGGGCTCTCCCTACAGG 473
Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
443 caccagagggc--accacaagtctctctgtggccaaagatgtgggcaaaagtgggcaaaa 500
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
474 AACCAGAGGCGGTATACAGTTCCTCTTGGCCAGGAGTGTGGGCGAGGGGTGGCGAGTG 533
Qy | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
501 gtggcaatctctctgctcctagctcaggtcaggtcaggtcaggtcaggtcaggtcaggt 540
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
534 TTGGGGGTCTCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 573

RESULT 2
LOCUS BE113924 305 bp mRNA EST 13-JUN-2000
DEFINITION UI-R-BJ1-aw1-a-09-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
ACCESSION BE113924
VERSION BE113924.1 GI:8506029
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 305)
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL discovery
MEDLINE 9704477
COMMENT Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 1-295, >LI_RN#LINE/L1
Seq primer: M13 Forward
POLYA-No. Location/Qualifiers
1..305
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ1-aw1-a-09-0-UI"
/clone_lib="UI-R-BJ1"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
library is a subtracted library derived from the following
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
For a detailed description of the library from which this
clone was derived, please visit our web site at
rates.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)

FEATURES
source
1..305
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ1-aw1-a-09-0-UI"
/clone_lib="UI-R-BJ1"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
library is a subtracted library derived from the following
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
For a detailed description of the library from which this
clone was derived, please visit our web site at
rates.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)

BASE COUNT 81 a 97 c 85 g 42 t
ORIGIN
Query Match 14.1%; Score 154; DB 33; Length 305;
Best Local Similarity 76.7%; Pred. No. 5.8e-35;
Matches 227; Conservative 0; Mismatches 65; Indels 4; Gaps 3;
QY 306 ggtatgtcttccctctgaaggtctagctctccctccatcctcattggtgagggagc 365
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
295 GGCATGTCCGCTCCCTCTGAATGTTAGCTCTCCCTCCCTTGGGATTGAGTGCAGGAGC 236
QY 366 tgttgacaggtctctcctcaaatccggtgcagctctgacccg-caggtctctgagctg 424
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
235 TG-TTGACTGCGCTCTCAGATCAGGGGGTGTCTGAGGCCCGGGACCTGCTGCTG 177
QY 425 cctgtgcaatcttcccgcaaccagaggcacc--caagtctctcttggggcaagatgt 482
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
176 AGTGCCCATCTCTCTGTTCCAAAGGTCCCTTACAGTTTCTCTCTGGCCAGGATGT 117
QY 483 gggcaaggtgggcaagtggtgcaatctctctgctcctcagctcaggtgctccac 542
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
116 GGTAAAGGTGGCAATTATGGCAGTCTTCTCTCTGCTCAGTCTCAGTGTGCCACCT 57
QY 543 ttctggcaatccgctctctctcctcagaggttgggagcagggagctgtggccg 598
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
56 GTCTGGGCAPTGAGCTCTCTCTCCCAAGGGTTTGGGACACAGACCTGTGGGCGG 1

```

library is a subtracted library derived from the following tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, ventricle at 13 dpc, and adult heart. For a detailed description of the library from which this clone was derived, please visit our web site at [ratseq.ing.uiowa.edu](http://ratseq.ing.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG\_SEQ=None found"

BASE COUNT 81 a 96 c 86 g 42 t

Query Match 14.0%; Score 152.4; DB 33; Length 305;  
Best Local Similarity 76.4%; Pred. No. 1.7e-34;  
Matches 226; Conservative 0; Mismatches 66; Indels 4; Gaps 3;  
QY 306 ggtatgtttccctctgaaggtctagctctccctccatggatgggtgcaggagc 365  
DB 295 GGCATGTCGCTCCCTCTGAATGTTTAGCTCCCTCCCTGGGATTTGAGTGCAGGAGC 236  
QY 366 ttttgaccaggtctctcaaatcgggtgagctctggaccg-caggctcctgtagcttg 424  
DB 235 TG-TTGACTGGGTCCCTTCAGATCAGGCGGTGCTCTGGAGCGCGGGACCTGCTGCTTG 177  
QY 425 cctgtgcaatcttccgcaccagagggcacc--caagttctcttgggccaaggtgt 482  
DB 176 AGTGCCCATCTTCTCTCCCAAGGTCCCTTACAGTTTCTCTGGGCGGAGGATGT 117  
QY 483 gggcaaggtgggagagtggaatctctctcctcctagctcaggtatggcctcac 542  
DB 116 GGGTAAGGTTGGGATATTGGCAGTCTTCTCTCTGCTGCTCAGTCTCAGGTGTTCCACCT 57  
QY 543 tctgggcaatccctctctctccacaggtttgggagcaggaggtgtgggccc 598  
DB 56 GTCGGGCAATGAGCTCTCTCCCAAGGGGTTTGGGAGCAGAGACCTGTGGCCCG 1

RESULT 4  
LOCUS C07078 493 bp mRNA EST 23-AUG-1996  
DEFINITION C07078 Rat pancreatic islet cDNA Rattus norvegicus CDNA, mRNA sequence.  
ACCESSION C07078  
VERSION C07078.1 GI:1503854  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 493)  
Takeda, J.  
AUTHORS Large scale collection of expressed sequence tags (ESTs) from rat pancreatic islet cDNA library  
TITLE Unpublished (1996)  
JOURNAL Contact: Jun Takeda  
COMMENT Institute for Molecular and Cellular Regulation, Gunma University 3-39-15 Showa-machi, Maebashi Gunma 371, Japan  
Tel: 272-20-8856  
Fax: 272-20-8896  
Email: jtakeda@sb.gunma-u.ac.jp.  
Location/Qualifiers  
1. .493  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone\_lib="Rat pancreatic islet cDNA"  
/tissue\_type="pancreatic islet"  
/notes="Vector: Lambda ZAPII; Site: 1: EcoRI; Site 2: XhoI; mRNA was prepared from normal rat islets. cDNA was directionally synthesized from the Xho I in the vector to the EcoRI site"

FEATURES  
source

BASE COUNT 73 a 134 c 148 g 129 t 9 others

Query Match 13.3%; Score 144.8; DB 36; Length 493;  
Best Local Similarity 62.9%; Pred. No. 3.7e-32;  
Matches 300; Conservative 0; Mismatches 166; Indels 11; Gaps 5;

QY 164 tcaagctgtcctctgtgaggaaccagagggcctgtcctactctactctg--gcccc 221  
DB 2 TCGAGCTCTCTGGCGGCGGAGCAACAGCAAGATCTGCCTCAGTTCTCTGGGAGCTTC 61  
QY 222 taagcacaggggcttagatggtgctaggtgttttctctctagagcctgaatgtgg--gc 279  
DB 62 AGTGCCACAGAGTTCAGATGGCGTTTGGTGTTCCTCTGGCGTCTGANAATGTGTGTGC 121  
QY 280 aagagtagtctctctgttctcctaggtatgtcttccctctgaagctctagctctccc 339  
DB 122 AGANTGCAGTCTCTCTGGTTTCCAGGGGTGTCTGCTCTGANGGTGTAGCTCTCCC 181  
QY 340 ttccatggga--atgggtgcaggagagctgtt----tgaccaggtcctctcaaatccgggtg 395  
DB 182 TCCCACAGGATTTGGGTGCAGAGAACTGTTTATCTGGTCTGATTCTTCAGGTTCTGGCG 241  
QY 396 cagctctggacgcag-gctcctgtagcttgcctgtgcaatcttcccccacccagagc- 453  
DB 242 GTGCTCTACACACAGCGGTCTGCAGCTCCTGGGCGCTCCCTCAGCGGAACCCAGANGCC 301  
QY 454 -accacagtttctctctgggcaaggtgtgggcaaggtgggagagtggaatctct 512  
DB 302 TTATACAGTTCCTCTTGGGCCANGATGTGGCCANGGGTGGCTGTATTGGTGTCTCT 361  
QY 513 cctgccttagctctcaggtatggcctcactctgtggcaatccgctctctctccacagg 572  
DB 362 TCCGCTCTGCACGCTCAGGAGTGCACCTGACCCANGGGTGGCTCTCTCTCAAGTT 421  
QY 573 gtttgggagcgggagctgtggccggtatcaggcaaggtttgaggcaacagtta 629  
DB 422 GTCTGGGANCANAAATGCTGTGGGCGGGATCGCGCGGTGTGGTCTCCCGGTAA 478

RESULT 5  
LOCUS H33201 368 bp mRNA EST 02-APR-1998  
DEFINITION EST108975 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA clone RPNAJ24 3' end, mRNA sequence.  
ACCESSION H33201  
VERSION H33201.1 GI:978618  
KEYWORDS EST.  
SOURCE Rattus sp.  
ORGANISM Rattus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 368)  
Lee, N.H.; Weinstein, K.G.; Kirkness, E.F.; Earle-Hughes, J.A.; Fuldner, R.A.; Marmaras, S.; Glodet, A.; Gocayne, J.D.; Adams, M.D.; Kerlavage, A.R.; Fraser, C.M. and Venter, J.C.  
AUTHORS Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment  
TITLE Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)  
JOURNAL 95396786  
MEDLINE Other ESTs: EST108976  
COMMENT Contact: Lee, NH  
ATCC  
The Institute for Genomic Research  
9712, Medical Center Drive, Rockville, MD 20850, USA  
Tel: (301)-838-3529  
Fax: (301)-838-0208  
Email: nhlee@tigr.org  
For clone availability please contact the TIGR Database (tdbinfo@tigr.org)



Seq primer: M13 - 21.  
Location/Qualifiers  
1. .368  
/organism="Rattus sp."  
/db\_xref="ATCC (inhost):2003480"  
/db\_xref="taxon:10118"  
/clone="RPNJ24"  
/clone.lib="Rat PC-12 cells, NGF-treated (9 days)"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; poly(A)+ RNA was purified from 9-day NGF treated PC12 cells. cDNA was constructed using an oligo-dT primer and directionally cloned using the Lambda ZAP II Vector Kit by Stratagene"

BASE COUNT 80 a 77 c 90 g 119 t 2 others  
ORIGIN

Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .353  
/organism="Rattus sp."  
/db\_xref="ATCC (inhost):2003029"  
/db\_xref="taxon:10118"  
/clone="RPCCL41"  
/clone.lib="Rat PC-12 cells, untreated"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; poly(A)+ RNA was purified from untreated PC12 cells cultured for 9 days. cDNA was constructed using an oligo-dT primer and directionally cloned using the Lambda ZAP II Vector Kit by Stratagene"

BASE COUNT 44 a 98 c 107 g 98 t 6 others  
ORIGIN

Query Match 11.4%; Score 124.6; DB 37; Length 353;  
Best Local Similarity 65.3%; Pred. NO. 3.5e-26;  
Matches 213; Conservative 0; Mismatches 104; Indels 9; Gaps 2;  
QY 202 cctacttctactgggcccctacgacaggggcccctagatggtgcttagtcttctcct 261  
Db 4 CCGCCCTTTCTGGGAGCTTCCTGTCACAGGGGTCCAGATGGCGTTGGTGTTCCTCT 53  
QY 262 agagcctgaaatgggacagagtagtctcctctggttctcttagtcttagtctccctc 321  
Db 64 GGTGTGAGAGATGTGGGACAGTGCAGTCTCTCTGTTTCCAGGCGGTGTGCTCTIN 123  
QY 322 tgaaggtctagctctcctccatgggatatgggtgcagggagctgtttgaccaggtcct 381  
Db 124 TAAAGGTTTACGCTCCCTCCACGGGATTTGGGTGCAGAGAACTGTTATCCCG--TCT 181  
QY 382 ctcaaatccgggtgagcttgagccagcagctccttagcttgcctgctgcaatctccc 441  
Db 182 GTTCTCTACAGTTCGGCGGTGTCTCAGCAGAGTTCCTGCGCGTCTCTGGGCCCTCCCC 241  
QY 442 -----gcacccagagggcaccacaaagtctctctgggccaagatgtggcgaagtgg 494  
Db 242 CACGGGAGCCGAGAGGCTTATACAGTTTCTCTTGGGCCAGGATGTGGCANGGGTGG 301  
QY 495 gcagaagtggcaatctctcctgcct 520  
Db 302 GCANTGTTGGTGTCTCTCTGCTCT 327

Query Match 13.1%; Score 142.4; DB 37; Length 368;  
Best Local Similarity 74.7%; Pred. NO. 1.7e-31;  
Matches 195; Conservative 0; Mismatches 57; Indels 9; Gaps 1;  
QY 220 cctacgcacagggccctagatggtgctaggtgttttctcttagagcctgaaatgtgggc 279  
Db 3 CCTGTGCACAGGGGCCAGATGGCACTAGGCAATTNCTCTAGAGTCAGAAATGTGGGC 62  
QY 280 agagtagtctctctgttctcttagtctctcctcctcctcctcctcctcctcctcctc 339  
Db 63 AGAGAGTAGTCTCTCTGCTTCCAGGTTATCTGCCCCCTCTGAAGTCTAGCACTCCC 122  
QY 340 ttccatgggatatgggtcaggagccttttgaccagctctcctcaaatccgggtgcagt 399  
Db 123 TCCACAGGATTTGGGTGAGGGAAGTGTTCACCAAGTCCCTTCAGATGAGGTGGTGT 182  
QY 400 ctggaccgcaggtcctcttagcttgcctgctgctcaatcttccgcacccagggcaccocaa 459  
Db 183 CTAGACTGAG-----CATTGAGTCCCTATGTTCTTCGTTCCAGAGCCCTATA 233  
QY 460 gttctcttgggccaagat 480  
Db 234 GGTATTATTTCTAATGAGGAT 254

RESULT 7  
AW921512 505 bp mRNA EST 25-MAY-2000  
LOCUS EST352816 Rat gene index, normalized rat, norvegicus, Benco Soares  
DEFINITION Rattus norvegicus cDNA clone RGIIA14 5' end, mRNA sequence.  
ACCESSION AW921512  
VERSION AW921512.1 GI:8087336  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 505)  
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,  
Kerlavage,A.R. and Adams,M.D.  
TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat  
Gene Index  
JOURNAL Unpublished (1998)  
COMMENT Contact: Lee, NH  
ATCC  
The Institute for Genomic Research  
9712, Medical Center Drive, Rockville, MD 20850, USA  
Tel: (301)-838-3529  
Fax: (301)-838-0208  
Email: nhlee@tigr.org  
This clone is available through the ATCC, contact the ATCC

RESULT 6  
H32857 353 bp mRNA EST 02-APR-1998  
LOCUS EST108352 Rat PC-12 cells, untreated Rattus sp. cDNA clone RPCCL41  
DEFINITION 5' end, mRNA sequence.  
ACCESSION H32857  
VERSION H32857.1 GI:978274  
KEYWORDS EST.  
SOURCE Rattus sp.  
ORGANISM Rattus sp.  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 353)  
AUTHORS Lee,N.H., Weinstein,K.G., Kirkness,E.F., Earle-Hughes,J.A., Fuldner  
,R.A., Marmaras,S., Glodek,A., Gocayne,J.D., Adams,M.D., Kerlavage  
,A.R., Fraser,C.M. and Venter,J.C.  
TITLE Comparative expressed-sequence-tag analysis of differential gene  
expression profiles in PC-12 cells before and after nerve growth  
factor treatment  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)  
MEDLINE 95396786  
COMMENT Contact: Lee, NH  
ATCC  
The Institute for Genomic Research  
9712, Medical Center Drive, Rockville, MD 20850, USA  
Tel: (301)-838-3529  
Fax: (301)-838-0208  
Email: nhlee@tigr.org  
For clone availability please contact the TIGR Database  
(tdbinfo@tdb.tigr.org)

tel#703-365-2700 for further information  
Seq primer: M13 Reverse

## FEATURES

source

Location/Qualifiers

1. .505

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

/clone="RG1A14"

/bento\_lib="Rat gene index, normalized rat, norvegicus,

liver, embryo, heart, muscle, spleen"

/tissue\_type="mix - brain, ovary, placenta, kidney, lung,

/lab\_host="SOLR"

/note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:

XhoI; Estimated insert size approx. 1 kb"

72 a 123 c 149 g 161 t

BASE COUNT

ORIGIN

Query Match 10.9%; Score 118.4; DB 25; Length 505;  
Best Local Similarity 68.3%; Pred. No. 2.7e-24;  
Matches 164; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 145 tttctgtccactgacttcaagctgtctgtgtgaggaacacgaagggcctgtccc 204

Db 186 TGCTATGGGGTGTGCTTATGAGCTCTCCGCGGGGGGAGCAACACGAGAGATGTGCGCT 245

QY 205 tacttactggccctacgcacaggggcttagatgtgctaggtgtttctctctaga 264

Db 246 GCCCTTTCGGGAGCTTCCTCGTCACACAGGTCACAGATGGCATTTGTTTCTCTGGA 305

QY 265 gctgaatgtgggagagtagtctctctgttctctaggtatgtcttccctctga 324

Db 306 ATCAGTAATGTGGGAGAGTGCTCTCTCTGTTTCCAGGTGTGCTGCCCTCTGTA 365

QY 325 aggtctagctctctctccatgggagtaggtgagggagctgtttgacaggtctctc 384

Db 366 AGGTTAGCTCTCCCTCTCACGGGATTGGGTGCAGAGACTGTTTATCTGGTCCGTC 425

RESULT 8

A1237080

LOCUS

DEFINITION EST 31-JAN-1999

ESN233642 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone

ROVDO45 3' end, mRNA sequence.

ACCESSION A1237080

VERSION A1237080.1 GI:3830586

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1998)

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13-21.

Location/Qualifiers

1. .493

/organism="Rattus sp."

/db\_xref="ATCC (inhost):2042239"

/clone="RG1A14"

/bento\_lib="Rat gene index, normalized rat, norvegicus,

liver, embryo, heart, muscle, spleen"

/tissue\_type="mix - brain, ovary, placenta, kidney, lung,

/lab\_host="SOLR"

/note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:

XhoI; Estimated insert size approx. 1 kb"

/note="Organ: ovary; Vector: pT73pac; Site\_1: EcoRI;  
Site\_2: NotI"

BASE COUNT 86 a 110 c 133 g 163 t 1 others  
ORIGIN

Query Match 10.5%; Score 114.8; DB 9; Length 493;  
Best Local Similarity 71.8%; Pred. No. 3.2e-23;  
Matches 163; Conservative 0; Mismatches 63; Indels 1; Gaps 1;

QY 145 tttctgtccactgacttcaagctgtctgtgtgaggaacacgaagggcctgtccc 204

Db 242 TGCTATGGGGTGTGCTTATGAGATCTCCAGGAGGAGTAACACGAGGAGGACCTGCACC 301

QY 205 tacttactgttggccctacgcacaggggcttagatgtgctaggtgtttctctctaga 263

Db 302 GCCCTTTCGAAAGTCCCGCTGCACACAGAGTCCCAATGGCTTTGTTTCTCTG 361

QY 264 agctgaatgtgggagagtagtctctctgttctctaggtatgttctctctctctga 323

Db 362 AGTCAGAAATGTGGGAGAGGTAGTCTCTCTGTTTCCAGGATGTCTGCTCTCTG 421

QY 324 aggtctagctctctccatgggagtaggtgagggagctgtttctctctctaga 370

Db 422 AAGTTAGCTCTCCCTCTCACGGGATTGGGTGCAGAGAAATGTTT 468

RESULT 9

A1555421

LOCUS

DEFINITION

UI-R-C2p-qp-f-08-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone

UI-R-C2p-qp-f-08-0-UI 3', mRNA sequence.

ACCESSION A1555421

VERSION A1555421.1 GI:4487784

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

1 (bases 1 to 254)

Bonald, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319.335.8250

Fax: 319.335.9565

Email: msoares@blue.weeg.uiowa.edu

cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone

distribution: clones will be available through Research Genetics

(www.resgen.com) The following repetitive elements were found in

this cDNA sequence: 8-246, &gt;U1\_RN#LINE/11

Seq primer: M13 Forward.

Location/Qualifiers

1. .254

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-C2p-qp-f-08-0-UI"

/bento\_lib="UI-R-C2p"

/dev\_stage="adult"

/lab\_host="DHI08 (Life Technologies)"

/note="vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-C2p

library is a subtracted library derived from the UI-R-C1

library, which is a subtracted library derived from the

UI-R-C0 library. The UI-R-C0 library consisted of a

mixture of individually tagged normalized libraries





REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Rattus.  
1 (bases 1 to 460)  
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,  
Kerlavage,A.R. and Adams,M.D.  
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat  
Gene Index  
Unpublished (1998)  
Contact: Lee, NH  
ATCC

The Institute for Genomic Research  
9712, Medical Center Drive, Rockville, MD 20850, USA  
Tel: (301)-838-3529  
Fax: (301)-838-0208  
Email: nhlee@tigr.org  
This clone is available through the ATCC, contact the ATCC  
tel#703-365-2700 for further information  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. 460  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="RGH14"  
/clone\_lib="Rat gene index, normalized rat, norvegicus,  
Bento Soares"  
/tissue.type="mix - brain, ovary, placenta, kidney, lung,  
liver, embryo, heart, muscle, spleen"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Estimated insert size approx.1 Kb"  
XhoI; Estimated insert size approx.1 Kb"  
85 a 123 c 134 g 118 t

BASE COUNT  
ORIGIN

Query Match 8.6%; Score 93.2; DB 25; Length 460;  
Best Local Similarity 63.0%; Pred. No. 8.5e-17;  
Matches 160; Conservative 0; Mismatches 93; Indels 1; Gaps 1;

QY 502 tggcaatctctcgtccctagcgtctcagagattggtccctcactctgggcaatccgctctc 561  
Db 9 TGGTGGTCTCTCCGCTCTCCAGCTTCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGC 68  
QY 562 tcttccacaggtttgggagcaggagctgtggtggtggtggtggtggtggtggtggtggtggt 621  
Db 69 TTTCCCAAGTGTCTGGGAGCAGAGAGCTGCTGCGGGCGGGATCGCGGGATGGGAC 128  
QY 622 accagttagaactggaagtgtcaggtccagaggaatttgcttctgtgtctgtcgtgagt 681  
Db 129 TCCGGCAA-ACACAGGAGCTGCCGGTCTCTAGATGACTCTGCTCTGTGTGTGTCGATC 187  
QY 682 ccaccagcaggtcacttgagcagcaaaaattgtttccctcgtctcaggtcaggtcaggt 741  
Db 188 TCACAGCAGCTCTCTTGAGCAGAGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 247  
QY 742 ttgcacctcagggt 755  
Db 248 TTTGCTCGCGGGT 261

RESULT 14  
A1137948/c  
LOCUS  
DEFINITION  
UI-R-CO-hj-b-12-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone  
UI-R-CO-hj-b-12-0-UI 3', mRNA sequence.  
ACCESSION  
A1137948  
VERSION  
A1137948.1 GI:3638725  
KEYWORDS  
EST.  
SOURCE  
Norway rat.  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1 (bases 1 to 507)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Oligo-dt track not found. Not 1 site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.  
Fatima Pinaldo, Ph.D. Clone Distribution: clones will be available  
through Research Genetics This clone is also available through the  
I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE  
ID-1781490 The following repetitive elements were found in this  
cDNA sequence: 9-97, >RMR34LTR 173-407, >LL\_RNLINE/L1  
Seq primer: M13 Forward  
POLYA-NC.

Location/Qualifiers  
1. 507  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CO-hj-b-12-0-UI"  
/clone\_lib="UI-R-CO"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CO  
library is a subtracted library derived from the UI-R-Al  
and UI-R-E1 libraries. The UI-R-Al library consisted of a  
mixture of individually tagged normalized libraries  
constructed from rat placenta, adult lung, brain, liver,  
kidney, heart, spleen, ovary, and muscle. The UI-R-E1  
library consisted of a mixture of individually tagged  
normalized libraries constructed from 8, 12 and 18-day  
embryo. The tag is a string of 3-5 nucleotides present  
between the Not I site and the oligo-dt track which  
allows identification of the library of origin of a clone  
within the mixture. The subtracted library (UI-R-CO) was  
constructed as follows: PCR amplified cDNA inserts from a  
pool of UI-R-Al and UI-R-E1 clones from which 3' ESTs had  
been derived was used as a driver in a hybridization with  
the pooled UI-R-Al and UI-R-E1 library in the form of  
single-stranded circles. The remaining single-stranded  
circles (subtracted library) was purified by  
hydroxyapatite column chromatography, converted to  
double-stranded circles and electroporated into DH10B  
bacteria (Life Technologies) to generate the UI-R-CO  
library. This procedure has been previously described  
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,  
1996)"

BASE COUNT 150 a 124 c 133 g 98 t 2 others  
ORIGIN

Query Match 8.4%; Score 91.8; DB 8; Length 507;  
Best Local Similarity 70.3%; Pred. No. 2.3e-16;  
Matches 137; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

QY 161 ctttcaagctgtctctgtgtgcaggaaccagagggcctgtccctacttctactgtg-gcc 219  
Db 363 CCTACGAGCTCTCTGTGTCGGCAGCAACAGGAGATCTGGCTGCCGTTTCCAGAGGCC 304  
QY 220 cctacgacagggggccctagatgt 279  
Db 303 TCCATGTACAGGGTTCAGATGGT 244  
QY 280 acagatgtctcctctgt 339  
Db 243 AGAGTGCAGCTCTTCTGT 184

QY 340 ttccatgggatagg 354  
 Db 183 TCCCGGGATTCTAG 169

## RESULT 15

AII76946

LOCUS

DEFINITION AII76946 457 bp mRNA EST 08-JAN-1999  
 EST220552 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone  
 ROVEY92.3' end, mRNA sequence.

ACCESSION

AII76946

VERSION

AII76946.1

KEYWORDS

EST.

SOURCE

Rattus sp.

ORGANISM

Rattus sp.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1998)

On Oct 8, 1998 this sequence version replaced gi:3727584.

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13-21.

FEATURES

source

1..457

Location/Qualifiers

/organism="Rattus sp."

/db\_xref="taxon:10118"

/clone="ROBY92"

/clone\_lib="Normalized rat ovary, Bento Soares"

/note="Organ: ovary; Vector: pT73Pac; Site\_1: EcoRI;

Site\_2: NotI"

BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 7.6%; Score 83.2; DB 9; Length 457;

Matches 122; Conservative 0; Mismatches 28; Indels 2; Gaps 2;

QY 16 cttgtttccattctgcacagtggtttgac-cttctatcacgctgtgtgcaggagtg 74  
 Db 80 CCTGTTCGTATTTCGACAGTGGTTAGACTGTCTTAAAGCTGTGTTCAAGAGTGC 139

QY 75 tgtagacctattttctctgtttcttcagccaggttacagacagagtggttctactgtca 134  
 Db 140 TGTGACCTGTTTCCCGTGTTCCTTCAGCCAGTATGCGGACAGAGAGTGTCT-TTG 198

QY 135 gatgttagctgttctctgtccactgactttca 166  
 Db 199 GCGGTGTAGTTCCTTCCTATCTACAGGCTTCA 230

Search completed: November 3, 2000, 22:08:08  
 Job time: 26222 sec

US 091014230WP1



Creation date: 04-08-2003  
Indexing Officer: HNGUYEN28 - HAO NGUYEN  
Team: OIPEBackFileIndexing  
Dossier: 09101423

Legal Date: 05-03-2001

No.	Doccode	Number of pages
1	SRNT	90

Total number of pages: 90

Remarks:

Order of re-scan issued on .....

